What Works Clearinghouse™

INSTITUTE OF EDUCATION SCIENCES

Adolescent Literacy February 2013

LANGUAGE!®

Program Description[1]

LANGUAGE!® is a language arts intervention designed for struggling learners in grades 3–12 who score below the 40th percentile on standardized literacy tests. The curriculum integrates English literacy acquisition skills into a six-step lesson format. During a daily lesson, students work on six key literacy strands (which the developer calls "six steps from sound to text"): phonemic awareness and phonics (word decoding), word recognition and spelling (word encoding), vocabulary and morphology (word meaning), grammar and usage (understanding the form and function of words in context), listening and reading comprehension, and speaking and writing.

Research[2]

The What Works Clearinghouse (WWC) identified one study of *LANGUAGE!*® that both falls within the scope of the Adolescent Literacy topic area and meets WWC evidence standards. The one study meets standards with reservations and includes 1,272 students in grades 9 and 10 in one school district in Florida.

The WWC considers the extent of evidence for *LANGUAGE!*® on the literacy skills of adolescent readers to be small for two domains: reading fluency and comprehension. Two other domains are not reported in this intervention report. (See the Effectiveness Summary on p. 4 for further description of all domains.)

Effectiveness

LANGUAGE!® was found to have no discernible effects on both reading fluency and comprehension for adolescent readers.

Table 1. Summary of findings[3]

Outcome domain	Rating of effectiveness	Improvement index *(percentile points)*		Number of studies	Number of students	Extent of evidence
		Average	Range			
Reading fluency	No discernible effects	0	na	1	640	Small
Comprehension	No discernible effects	–5	na	1	632	Small

na = not applicable

Report Contents

Overview	p. 1
Program Information	p. 2
Research Summary	p. 3
Effectiveness Summary	p. 4
References	p. 5
Research Details for Each Study	p. 7
Outcome Measures for Each Domain	p. 10
Findings Included in the Rating for Each Outcome Domain	p. 11
Endnotes	p. 12
Rating Criteria	p. 13
Glossary of Terms	p. 14

Program Information

Background

LANGUAGE!® is distributed by Voyager Learning, a division of Cambium Learning Group, Inc. Address: 17855 Dallas Parkway, Suite 400, Dallas, TX 75287. Email: customersupportvel@voyagerlearning.com. Web: www.voyagerlearning.com. Telephone: (888) 399-1995.

Program details

LANGUAGE!® is designed for students in grades 3–12 who score below the 40th percentile on standardized literacy tests. The curriculum includes six levels, A–F, each with six units of instruction and 10 lessons per unit. Students enter the curriculum at skill level A, C, or E, based on a group-administered placement test. Students demonstrating a deficiency in basic decoding start the program at Level A. Students showing proficiency with beginning sound/symbol correspondences but deficiencies at higher levels of word analysis start the program at Level C. Students in grades 7–12 who show proficiency with sound/symbol correspondences and higher levels of word analysis start the program at Level E.

In a typical 90-minute lesson, time is distributed across the six literacy strands. When more time is available, additional instructional options are possible, such as listening to complex text selections (which the developer calls "Challenge Text"), answering critical-thinking questions through group discussions, writing, and speaking activities.

The program also includes an online tool called *VocabJourney*®, which is designed to provide students with additional opportunities to practice their vocabulary.

Cost

Individual materials for the *LANGUAGE!*® curriculum range in price. A teacher's set for each level costs $353 and includes teacher and student textbooks, online technology applications including *VocabJourney*®, data management, and instructional and assessment tools. A student set costs $69 and includes student text, assessment materials, access to *VocabJourney*®, and other online tools (such as *eReader*). Prices are effective December 2012. Additional materials are available at additional cost and can be found on the distributor's website.

Research Summary

The WWC identified 16 studies on the effects of *LANGUAGE!*® on the literacy skills of adolescent readers. The WWC reviewed seven of those studies against group design evidence standards. One study (Zmach, Chan, Salinger, Chinen, Tanenbaum, & Taylor, 2009) is a quasi-experimental design that meets WWC evidence standards with reservations. This study is summarized in this report. Six studies do not meet WWC evidence standards. The remaining nine studies do not meet WWC eligibility screens for review in this topic area. Citations for all 16 studies are in the References section, which begins on p. 5.

Table 2. Scope of reviewed research

Grade	9, 10
Delivery method	Whole class
Program type	Curriculum

Summary of studies meeting WWC evidence standards without reservations

No studies of *LANGUAGE!*® met WWC evidence standards without reservations.

Summary of study meeting WWC evidence standards with reservations

Zmach et al. (2009) conducted a quasi-experimental study of middle and high school students in the Miami–Dade County Public School district. The researchers used a two-stage matching process to first select the study schools and then select the student samples from these schools. To identify intervention schools, the district selected Title I schools that already used or were going to use the *LANGUAGE!*® program in their Intensive Reading Plus (IR+) classes for struggling readers. The IR+ class is a 90-minute instructional block scheduled back to back with the regular English language arts class. Secondary school students who scored at Level 1 or 2 on the Florida Comprehensive Assessment Test (FCAT) and required intervention in decoding, fluency, vocabulary, and comprehension were eligible to enroll in an IR+ class; some schools used other factors, including performance on other reading assessments, prior enrollment in IR+ classes with no improvement in performance, and school staff recommendations, to determine eligibility for IR+ classes. To select the comparison schools, these intervention schools were first matched to schools based on school type, school size, and student demographics. Then, to select the comparison students, students in IR+ classes in the comparison schools were matched to intervention students using a propensity score matching process, based on pretest score, grade levels, and socio-demographic variables.

Although the study was conducted with students in grades 6–10, this intervention report presents results only for the high school students in grades 9–10.[4] The high school sample included eight intervention schools and ten comparison schools. The authors analyzed two student samples: one sample that had pretest scores for the Test of Silent Contextual Reading Fluency (TOSCRF), and one sample that had pretest scores for the FCAT. As no information on the extent of overlap between the two analytic samples was provided in the study, the WWC review process treats them as two distinct samples.

The TOSCRF student sample included 320 students who used the *LANGUAGE!*® curriculum and 320 students who used the district's regular reading curriculum in the IR+ classes.[5] The FCAT student sample included 316 students who used the *LANGUAGE!*® curriculum and 316 students who used the district's regular reading curriculum in the IR+ classes. The study reported student outcomes after one academic year of program implementation.

Effectiveness Summary

The WWC review of interventions for Adolescent Literacy addresses student outcomes in four domains: alphabetics, reading fluency, comprehension, and general literacy achievement. The one study that met WWC evidence standards reported findings in two of the four domains: (a) reading fluency and (b) comprehension. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *LANGUAGE!*® on adolescent readers for each domain. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 13.

Summary of effectiveness for the reading fluency domain

One study reported findings in the reading fluency domain.

Zmach et al. (2009) reported a negative, but not statistically significant, difference between the *LANGUAGE!*® group and the comparison group on the TOSCRF test. The effect size reported by the study authors was not large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). The WWC characterizes this study finding as an indeterminate effect.

Thus, for the reading fluency domain, one study showed indeterminate effects. This results in a rating of no discernible effects, with a small extent of evidence.

Table 3. Rating of effectiveness and extent of evidence for the reading fluency domain

Rating of effectiveness	Criteria met
No discernible effects *No affirmative evidence of effects.*	In the one study that reported findings, the estimated impact of the intervention on outcomes in the *reading fluency* domain was neither statistically significant nor large enough to be substantively important.
Extent of evidence	**Criteria met**
Small	One study that included 640 students in 18 schools reported evidence of effectiveness in the *reading fluency* domain.

Summary of effectiveness for the comprehension domain

One study reported findings in the comprehension domain.

Zmach et al. (2009) reported a negative, but not statistically significant, difference between the *LANGUAGE!*® group and the comparison group on the FCAT reading test. The effect size reported by the study authors was not large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). The WWC characterizes this study finding as an indeterminate effect.

Thus, for the comprehension domain, one study showed indeterminate effects. This results in a rating of no discernible effects, with a small extent of evidence.

Table 4. Rating of effectiveness and extent of evidence for the comprehension domain

Rating of effectiveness	Criteria met
No discernible effects *No affirmative evidence of effects.*	In the one study that reported findings, the estimated impact of the intervention on outcomes in the *comprehension* domain was neither statistically significant nor large enough to be substantively important.
Extent of evidence	**Criteria met**
Small	One study that included 632 students in 18 schools reported evidence of effectiveness in the *comprehension* domain.

References

Studies that meet WWC evidence standards without reservations

None

Study that meets WWC evidence standards with reservations

Zmach, C. C., Chan, T., Salinger, T., Chinen, M. H., Tanenbaum, C. T., & Taylor, T. S. (2009). *Evaluation of LANGUAGE! in Miami-Dade County Public Schools: Final report.* Washington, DC: American Institutes for Research.

Studies that do not meet WWC evidence standards

Cambium Learning Group. (n.d.). Elk Grove Unified School District, California. In *LANGUAGE! research base* (pp. 28–31). Longmont, CO: Author. 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.

Cambium Learning Group. (n.d.). Los Angeles County, California. In *LANGUAGE! research base* (pp. 50–51). Longmont, CO: Author. 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.

Cambium Learning Group. (n.d.). Systemwide results with LANGUAGE! curriculum 4th edition. In *LANGUAGE! research base* (pp. 72–73). Longmont, CO: Author. 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.

Greene, J. F. (1996). LANGUAGE! Effects of an individualized structured language curriculum for middle and high school students. *Annals of Dyslexia, 46,* 97–121. The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.

Sopris West Educational Services. (2006). *Caldwell County Schools, North Carolina: Upper elementary and middle school retrospective evaluation with state data.* Frederick, CO: Author. 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.

Additional source:

Cambium Learning Group. (n.d.). Caldwell County Schools, North Carolina. In *LANGUAGE! research base* (pp. 24–27). Longmont, CO: Author.

Sopris West Educational Services. (2006). *Lee County Public Schools, Florida: Districtwide middle and high school retrospective evaluation with state data.* Frederick, CO: Author. 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.

Additional source:

Cambium Learning Group. (n.d.). Lee County Public Schools, Florida. In *LANGUAGE! research base* (pp. 52–59). Longmont, CO: Author.

Studies that are ineligible for review using the Adolescent Literacy Evidence Review Protocol

Bergeson, T. (2004). *Grades 4–12 reading intervention materials review: Washington state evaluation report.* Olympia, WA: Washington State Department of Education. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Cambium Learning Group. (n.d.). Sacramento City Unified School District, California. In *LANGUAGE! research base* (pp. 68–71). Longmont, CO: Author. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Porter, R. P., & Clark, K. (2004). *Language and literacy for English learners, grades 7–12: Four programs of proven success.* Frederick, CO: Sopris West Educational Services. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample includes less than 50% general education students.

Robinson, C. (2002). *LANGUAGE!* Tallahassee, FL: Florida Center for Reading Research. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Sopris West Educational Services. (2005). *Elk Grove Unified School District, California: High school, middle school, and upper elementary school pilot evaluation.* Frederick, CO: Author. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Sopris West Educational Services. (2006). *Denver Public Schools, Colorado: Middle school pilot evaluation with students eligible for special education.* Frederick, CO: Author. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Sopris West Educational Services. (2006). *Great Falls Public Schools, Montana: Middle school retrospective evaluation.* Frederick, CO: Author. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Additional source:

Cambium Learning Group. (n.d.). Great Falls Public Schools, Montana. In *LANGUAGE! research base* (pp. 32–35). Longmont, CO: Author.

Sopris West Educational Services. (2007). *Hawthorne School District, California: Middle school pilot evaluation with students eligible for special education.* Frederick, CO: Author. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Additional source:

Cambium Learning Group. (n.d.). Hawthorne School District, California. In *LANGUAGE! research base* (pp. 36–39). Longmont, CO: Author.

Sopris West Educational Services. (2007). *Rialto Unified School District, California: Elementary school retrospective evaluation with state data.* Frederick, CO: Author. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Additional source:

Cambium Learning Group. (n.d.). Rialto Unified School District, California. In *LANGUAGE! research base* (pp. 62–67). Longmont, CO: Author.

Appendix A: Research details for Zmach et al., 2009

Zmach, C. C., Chan, T., Salinger, T., Chinen, M. H., Tanenbaum, C. T., & Taylor, T. S. (2009). *Evaluation of LANGUAGE! in Miami-Dade County Public Schools: Final report.* Washington, DC: American Institutes for Research.

Table A. Summary of findings

Meets WWC evidence standards with reservations

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
Reading fluency	640 students	0	No
Comprehension	632 students	–5	No

Setting

The student analysis sample included in this report is drawn from 18 high schools eligible for Title I funding in the Miami–Dade County Public School district.

Study sample

Selection of study schools. In this quasi-experimental study, researchers used a two-stage matching process to select the schools and student sample. The district selected nine Title I-eligible high schools and two Title I-eligible middle schools (grades 6–8) for the intervention group. Nine comparison high schools and two comparison middle schools were then matched with these intervention schools using the Euclidean distance approach, based on school type, school size, distribution of race/ethnic groups, percentage of students with limited English proficiency (LEP) status, and percentage of students who were eligible for free or reduced-price lunch in the 2005–06 school year. After student rosters indicated lower than expected student enrollment, one additional comparison high school was added to increase student sample size. One intervention high school was dropped from the study when its only *LANGUAGE!*® teacher left and could not be replaced, resulting in eight intervention high schools and ten comparison high schools. One middle school was dropped because of differences in characteristics from the other schools, resulting in two intervention middle schools and one comparison middle school.

Formation of two student analytic samples. The initial student sampling frame included all students in grades 6–10 in the study schools who were enrolled in an Intensive Reading Plus (IR+) class for struggling readers when the Test of Silent Contextual Reading Fluency (TOSCRF) was administered in the fall. The IR+ class is a 90-minute instructional block scheduled back to back with the regular English language arts class. Twenty percent of the eligible student sample did not have both the TOSCRF and Florida Comprehensive Assessment Test (FCAT) pretest scores that the authors had planned to use for the student matching process, so the authors created a TOSCRF student sample (comprising any student who had a TOSCRF score from that fall) and a FCAT student sample (comprising any student who had a FCAT score from the prior spring). A propensity matching model was used to match intervention students to students from the pooled comparison schools based on pretest score, grade levels, and socio-demographic variables.

Analysis sample. The ninth- and tenth-grade samples are the only samples which meet WWC evidence standards with reservations. Therefore, they are presented in this report. After dropping the unmatched students, students with missing information, and students who were no longer enrolled in an IR+ class in the same school in spring 2008, the TOSCRF and FCAT analytic high school samples included 640 students and 632 students, respectively. As no information on the extent of overlap between the two analytic samples was provided in the study, the WWC review process treats them as two distinct samples. The TOSCRF student sample included 320 students (190 ninth graders and 130 tenth graders) who used the *LANGUAGE!*® curriculum and 320 students (190 ninth graders and 130 tenth graders) who used the comparison curriculum in the IR+ classes.[6] The FCAT sample included 316 students (194 ninth graders and 122 tenth graders) who used the *LANGUAGE!*® curriculum and 316 students (194 ninth graders and 122 tenth graders) who used the comparison curriculum in the IR+ classes.

Characteristics of district and study schools. In 2006, 68% of ninth graders and 73% of tenth graders in the district scored below proficient (Level 2 or below) on the FCAT. During the 2005–06 school year:

- between 1,547 and 4,509 students attended the study high schools.
- the percentage of students in study high schools that were eligible for free or reduced-price lunch ranged from 40% to 68%.
- Black and Hispanic students represented between 77% and 99% of the student population in study high schools.

Characteristics of the TOSCRF student sample. Among the TOSCRF analytic sample:

- Forty-seven percent of ninth graders and 40% of tenth graders were female.
- Hispanic students represented 52% of ninth graders and 42% of tenth graders, while Black students represented 43% of ninth graders and 55% of tenth graders.
- About 68% of ninth graders and 59% of tenth graders were eligible for free or reduced-price lunch.
- Thirty-one percent of ninth graders and 33% of tenth graders were classified as receiving special education services.

Characteristics of the FCAT student sample. The demographic characteristics of the FCAT sample were similar to the TOSCRF sample. Among the FCAT student sample:

- Thirty-six percent of ninth graders and 48% of tenth graders were female.
- Hispanic students represented 52% of ninth graders and 39% of tenth graders, while Black students represented 45% of ninth graders and 56% of tenth graders.
- About 66% of ninth graders and 58% of tenth graders were eligible for free or reduced-price lunch.
- Thirty-six percent of ninth graders and 32% of tenth graders were classified as receiving special education services.

Intervention group

The intervention was delivered during the daily IR+ class, which typically lasted 90 minutes. The lessons were administered by the teacher to the whole classroom with some days set aside for differentiated instruction throughout the school year. Schools received a pacing guide designed to facilitate the completion of two book levels, out of six levels (book levels A-F described under program details), during the year. The intervention lasted a full academic year. The study team rated teachers on the fidelity of implementation of the curriculum. Fifty-four percent of teachers received a medium fidelity rating, and 46% of teachers received a low fidelity rating.

Comparison group

All comparison classrooms used the same commercially published curriculum in their daily IR+ class, which typically lasted 90 minutes (the study authors did not provide the name of the curriculum). The curriculum focused on strengthening reading and writing skills and developing vocabulary. Typically, 20 minutes of class time was spent on whole-group, direct instruction; 60 minutes was spent on small-group rotations; and 10 minutes was spent on whole-group, wrap-up instruction. The small-group rotations took the form of small-group direct instruction, technology-based individualized instruction, and modeled and independent reading.

Outcomes and measurement

The outcomes are gain scores for the TOSCRF and FCAT reading scores, i.e., gains in the TOSCRF and FCAT reading scores from the prior administration of the test to the current test. The TOSCRF was administered to the participating classes by study staff in October and then again seven months later. The FCAT is a state-wide assessment given each spring. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

Sopris West, the program publisher at the time of the study, was contracted to provide professional development to support teachers' implementation of the *LANGUAGE!*® curriculum. The intervention teachers attended a two-day training session before the school year. During the school year, teachers, coaches, and mentors received school visits from *LANGUAGE!*® trainers and National Trainers, who conducted classroom observations, provided individual coaching and professional development for teachers, modeled lessons, and held question-and-answer sessions. Coaches and school administrators received a half-day of initial training and a day of training in the fall and spring. The comparison teachers received the usual professional development services provided by their schools.

Appendix B: Outcome measures for each domain

Reading fluency	
Test of Silent Contextual Reading Fluency (TOSCRF)	This paper-and-pencil test measures silent reading fluency by having students separate words in passages presented entirely in upper case letters without punctuation or spaces between words. Researchers administered the test in the fall and spring of the school year and took about 15 minutes to administer, including practice and test completion time. The TOSCRF standard scores, which measure growth across age and grade levels, were used to compute the gain in TOSCRF scores between fall and spring administrations. The mean and standard deviation of the distribution of the TOSCRF standard scores are 100 and 15, respectively (as cited in Zmach et al., 2009).
Comprehension	
Reading comprehension	
Florida Comprehensive Assessment Test (FCAT) Reading Developmental Scale Scores (DSS)	This statewide, standardized test is administered to students in grades 3–10 each spring to measure overall reading ability. The study used the DSS to compute the gain in FCAT scores between spring 2007 and spring 2008. The DSS is a vertical scale that enables tracking of student progress over time. DSS values range from 86 (grade 3) to 3008 (grade 10). There are five achievement levels for each grade, ranging from level 1, which represents an inadequate level of success with the test content, to level 5, which represents mastery of the content (as cited in Zmach et al., 2009).

Appendix C.1: Findings included in the rating for the reading fluency domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Zmach et al., 2009[a]								
TOSCRF	Grades 9–10	18 schools/ 640 students	nr (nr)	nr (11.3)	–0.06	–0.01	0	0.89
Domain average for reading fluency (Zmach et al., 2009)						**–0.01**	**0**	**Not statistically significant**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the average change expected for all students who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. The statistical significance of the study's domain average was determined by the WWC; a study is characterized as having indeterminate effects when the single or mean effect is neither statistically significant nor substantively important. nr = not reported. TOSCRF = Test of Silent Contextual Reading Fluency.

[a] For Zmach et al. (2009), no corrections for clustering or multiple comparisons were needed. The mean difference (regression coefficient), effect size, and *p*-value presented here were reported in the original study (Table H-2, Model 2). The outcome measure is the gain in the TOSCRF scores from fall 2007 to spring 2008. Analyses of student outcomes were conducted with a three-level (student-teacher-school) hierarchical linear model (HLM). The effect size is calculated by dividing the mean difference by the standard deviation (11.3) of the fall 2007 TOSCRF scores of the comparison group (as shown in the tables in the study). However, the study text is not entirely consistent with the tables. In particular, on p. 38, the study authors refer to a different mean difference (–0.08 [instead of –0.06]) and indicate that the effect size is based on the standard deviation (11.3) of the spring 2008 (not fall 2007) TOSCRF scores of the comparison group. We could not resolve the discrepancies between the results presented in the text and table, but the effect size, rating, and improvement index do not change if one uses the findings presented in the text instead of those presented in the table.

Appendix C.2: Findings included in the rating for the comprehension domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Zmach et al., 2009[a]								
FCAT Reading DSS	Grades 9–10	18 schools/ 632 students	nr (nr)	nr (229.7)	–30.52	–0.13	–5	0.23
Domain average for comprehension (Zmach et al., 2009)						**–0.13**	**–5**	**Not statistically significant**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the average change expected for all students who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. The statistical significance of the study's domain average was determined by the WWC; a study is characterized as having indeterminate effects when the single or mean effect is neither statistically significant nor substantively important. nr = not reported. FCAT = Florida Comprehensive Assessment Test. DSS = Developmental Scale Scores.

[a] For Zmach et al. (2009), no corrections for clustering or multiple comparisons were needed. The mean difference (regression coefficient), effect size, and *p*-value presented here were reported in the original study (Table H-1, Model 2). The outcome measure is the gain in the FCAT DSS from spring 2007 to spring 2008. Analyses of student outcomes were conducted with a three-level (student-teacher-school) HLM. The study calculated the effect size by dividing the mean difference by the standard deviation (229.7) of the spring 2008 FCAT scores of the comparison group.

Endnotes

[1] The descriptive information for this program was obtained from a publicly available source: the program's website (www.voyagerlearning.com, downloaded January 2012). The program's registered trademark name is *LANGUAGE! The Comprehensive Literacy Curriculum*®. The WWC requests that distributors review the program description sections for accuracy from their perspective. The program description was provided to the distributor in February 2012, and we incorporated feedback from the distributor. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search reflects documents publicly available by December 2011.

[2] The studies in this report were reviewed using the Evidence Standards from the WWC Procedures and Standards Handbook (version 2.1), along with those described in the Adolescent Literacy review protocol, version 2.1. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[3] For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 13. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies. The WWC reviews of evidence for Adolescent Literacy address student outcomes in four domains: alphabetics, reading fluency, comprehension, and general literacy achievement. The one study of *LANGUAGE!*® that meets WWC evidence standards reported findings in two of the four domains: reading fluency and comprehension.

[4] Among the middle school sample, grades 7 and 8 are ineligible for review under the Adolescent Literacy review protocol, version 2.1, because at least 50% of each sample includes students classified as receiving special education services. The grade 6 sample is not included in this review because it does not meet WWC standards due to a confounding factor (in particular, there is only one school in the middle school comparison group).

[5] The sample size of 640 reported for the TOSCRF analysis sample is based on the study text (pp. v, 15, and 38) and Table H-2. However, there is a discrepancy with Tables 3-3 and E-22, which indicate that the total sample size was 642 (382 ninth graders and 260 tenth graders). We could not resolve this discrepancy, but the effect size, rating, and improvement index do not change based on the data presented in Tables 3-3 or E-22.

[6] Although Tables 3-3 and E-2 refer to 191 students in the intervention group and 191 students in the comparison group for the ninth-grade sample, we reported 190 students in each of these groups to be consistent with the total sample size of 640 that was reported for the high school sample.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, February). *Adolescent Literacy intervention report: LANGUAGE!*®. Retrieved from http://whatworks.ed.gov.

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC evidence standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC evidence standards with reservations	A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples.

Criteria used to determine the rating of effectiveness for an intervention

Rating of effectiveness	Criteria
Positive effects	Two or more studies show statistically significant positive effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important negative effects.
Potentially positive effects	At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
Mixed effects	At least one study shows a statistically significant or substantively important positive effect AND at least one study shows 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, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
Potentially negative effects	One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
Negative effects	Two or more studies show statistically significant negative effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important positive effects.
No discernible effects	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

Criteria used to determine the extent of evidence for an intervention

Extent of evidence	Criteria
Medium to large	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
Small	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

Clustering adjustment If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The design of a study is the method by which intervention and comparison groups were assigned.

Domain A domain is a group of closely related outcomes.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.

Extent of evidence An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p. 13.

Improvement index Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from –50 to +50.

Multiple comparison adjustment When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.

Rating of effectiveness The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 13.

Single-case design A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

Standard deviation The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.

Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$).

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the WWC Procedures and Standards Handbook (version 2.1) for additional details.