

The Enhanced Reading Opportunities Study

Findings from the Second Year of Implementation

Executive Summary

The Enhanced Reading Opportunities Study

Findings from the Second Year of Implementation

Executive Summary

NOVEMBER 2008

William Corrin
Marie-Andrée Somers
James J. Kemple
Elizabeth Nelson
Susan Sepanik
MDRC

With

Terry Salinger
Courtney Tanenbaum
American Institutes for Research

Paul Strasberg, Project Officer
Institute of Education Sciences

NCEE 2009-4037
U.S. Department of Education



U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover Whitehurst
Director

National Center for Education Evaluation and Regional Assistance
Phoebe Cottingham
Commissioner

November 2008

This report was prepared for the National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, under contract no. ED-01-CO-0111/0001 with MDRC.

This report is in the public domain. Authorization to reproduce it in whole or in part is granted. While permission to reprint this publication is not necessary, the citation should read: Corrin, W., Somers, M.-A., Kemple, J., Nelson, E., and Sepanik, S. (2008). *The Enhanced Reading Opportunities Study: Findings from the Second Year of Implementation* (NCEE 2009-4036). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the report.

To order copies of this report,

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 1398, Jessup, MD 20794-1398.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.
- Fax your request to 301-470-1244 or order online at www.edpubs.org.

This report is also available on the IES website at <http://ncee.ed.gov>.

Alternate Formats

Upon request, this report is available in alternate formats, such as Braille, large print, audiotape, or computer diskette. For more information, call the Alternate Format Center at 202-205-8113.

Acknowledgments

This study represents a collaborative effort among the authors and the staff from the participating school districts and schools, the program developers, our colleagues at MDRC and American Institutes for Research (AIR), and Institute of Education Sciences (IES) staff. The study has benefited especially from the time, energy, and commitment put forth by staff in the participating school districts to implement the two literacy programs used in the Enhanced Reading Opportunities (ERO) study, to allow access to classrooms, and to respond to requests for data.

At the U.S. Department of Education, Paul Strasberg, Marsha Silverberg, Phoebe Cottingham, and Ricky Takai at the Institute of Education Sciences provided helpful support and guidance on the design and execution of the evaluation and in the development of the report. Braden Goetz and Valerie Randall-Walker at the Office of Elementary and Secondary Education provided invaluable support to the school districts in their efforts to implement the supplemental literacy programs and meet the demands of the evaluation.

The study's technical working group provided valuable insights on the evaluation design, data analysis, and early versions of the report. We thank Donna E. Alvermann, Donald L. Compton, Robinson Hollister, Mark W. Lipsey, Robert H. Meyer, Christopher Schatschneider, Timothy Shanahan, and Catherine Snow for their expertise and guidance.

The listed authors of this report represent only a small part of the team involved in this project. Linda Kuhn and the staff at Survey Research Management managed and conducted the follow-up testing and survey data collection effort.

At AIR, Nancy Lang, Suzannah Herrmann, Kathryn Drummond, and Courtney Zmach conducted site visits and conducted phone interviews. Christopher Doyle and Andrea Olinger coordinated data management and conducted phone interviews. Nancy Lang processed data and ensured the thoroughness of the fidelity ratings.

At MDRC, Edmond Wong assisted with data collection and provided programming and analysis support. Corinne Herlihy and Kristin Porter served as school district coordinators. Daniel Fallon oversaw the ordering of the literacy assessment. Shirley James and her staff entered data. Gordon Berlin, Alison Black, Howard Bloom, Fred Doolittle, Corinne Herlihy, John Hutchins, Robert Ivry, Janet Quint, and Pei Zhu provided substantive expertise through their thoughtful comments on, and reviews of, this report. Edmond Wong and Mario Flecha assisted with report production. Robert Weber edited the report, and Stephanie Cowell and Inna Kruglaya prepared it for publication.

The Authors

Disclosure of Potential Conflicts of Interest¹

The research team for this evaluation consists of a prime contractor, MDRC, Inc., of New York City, NY, and two subcontractors, American Institutes for Research (AIR) of Washington, DC, and Survey Research Management (SRM) Corporation of Boulder, CO. None of these organizations or their key staff has financial interests that could be affected by findings from the evaluation of the two supplemental literacy interventions considered in this report. No one on the eight-member Expert Advisory Panel, convened by the research team once a year to provide advice and guidance, has financial interests that could be affected by findings from the evaluation. One member of the Expert Advisory Panel, Dr. Timothy Shanahan of the University of Illinois at Chicago, participated only in an early (2005) panel meeting on the study design. Subsequent to that meeting, he developed a commercial literacy intervention targeted to striving middle-school readers that might either compete with or be used along with the two programs for high school students chosen and evaluated as part of the current study. Dr. Shanahan had no role in the selection of the study programs or in the analysis of evaluation data.

¹Contractors carrying out research and evaluation projects for IES frequently need to obtain expert advice and technical assistance from individuals and entities whose other professional work may not be entirely independent of or separable from the particular tasks they are carrying out for the IES contractor. Contractors endeavor not to put such individuals or entities in positions in which they could bias the analysis and reporting of results, and their potential conflicts of interest are disclosed.

Executive Summary

This report presents findings from the Enhanced Reading Opportunities (ERO) study — a demonstration and rigorous evaluation of two supplemental literacy programs that aim to improve the reading comprehension skills and school performance of struggling ninth-grade readers. The U.S. Department of Education’s (ED) Office of Elementary and Secondary Education (OESE)¹ is funding the implementation of these programs, and its Institute of Education Sciences (IES) is responsible for oversight of the evaluation. MDRC — a nonprofit, nonpartisan education and social policy research organization — is conducting the evaluation in partnership with the American Institutes for Research (AIR) and Survey Research Management (SRM).

The present report — the second of three — focuses on the second of two cohorts of ninth-grade students to participate in the study and discusses the impact that the two interventions had on these students’ reading comprehension skills through the end of their ninth-grade year. The report also describes the implementation of the programs during the second year of the study and provides an assessment of the overall fidelity with which the participating schools adhered to the program design as specified by the developers. While this report focuses primarily on implementation and impacts in the second year of the study, comparisons between the first and second year of the study are also provided.² The key findings discussed in the report include the following:

- **On average, across the 34 participating high schools, the supplemental literacy programs improved student reading comprehension test scores by 0.08 standard deviation. This represents a statistically significant improvement in students’ reading comprehension (p-value = 0.042).**
- **Seventy-seven percent of the students who enrolled in the ERO classes in the second year of the study were still reading at two or more years below grade level at the end of ninth grade, relative to the expected reading achievement of a nationally representative sample of ninth-grade students.³ One of the two interventions — Reading Apprenticeship Aca-**

¹The implementation was initially funded by the Office of Vocational and Adult Education (OVAE), but this role was later transferred to OESE.

²James J. Kemple, William Corrin, Elizabeth Nelson, Terry Salinger, Suzannah Herrmann, and Kathryn Drummond, *The Enhanced Reading Opportunities Study: Early Impacts and Implementation Findings*, NCEE 2008-4015 (Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2008).

³Forty percent of ninth-graders nationally would be expected to score at two or more years below grade level on the same assessment.

demie Literacy (RAAL) — had a positive and statistically significant impact on reading comprehension test scores (0.14 standard deviation; p-value = 0.015). Although not statistically significant, a positive impact on reading comprehension (0.02 standard deviation) was also produced by the other intervention, Xtreme Reading. The difference in impacts between the two programs is not statistically significant, and thus it cannot be concluded that RAAL had a different effect on reading comprehension than Xtreme Reading.⁴

- The overall impact of the ERO programs on reading comprehension test scores in the second year of implementation (0.08 standard deviation) is not statistically different from their impact in the first year of implementation (0.09 standard deviation), nor is each intervention’s impact in the second year of implementation statistically different from its impact in the first year.**
- The implementation fidelity of the ERO programs was more highly rated in the second year of the study than in the first year. In comparison with the first year, a greater number of schools in the second year of the study were deemed to have programs that were well aligned with the program developers’ specifications for implementation fidelity (26 schools in the second year, compared with 16 schools in the first year), and fewer schools were considered to be poorly aligned (one school in the second year, compared with 10 schools in the first year).**

⁴It is important to note that the ERO study is an evaluation of a class of reading interventions, as represented by Xtreme Reading and RAAL, as well as an evaluation of each of these two programs separately. The purpose of the study is not to test the differential impact of these two interventions; while Xtreme Reading and RAAL do differ in some respects, they are both full-year supplemental literacy courses targeted at struggling adolescent readers that share many common principles, and hence there was no prior expectation that they would produce substantially different impacts. As noted below, the design of the study is such that programs are randomized to schools; however, the purpose of this randomization was to ensure that each program developer was assigned a fair draw of schools in which to implement its program, rather than to test for a differential impact between the two interventions. By this token, the statistical model chosen for the impact analysis does not utilize the school-level randomization feature of the research design; nor is the sample size large enough to detect policy-relevant differences in impacts across the two programs. Because Xtreme Reading and RAAL represent the same type of intervention, this study was designed to test their joint or overall impact. Statistical tests were used to confirm that the difference in impacts between the two programs is not statistically significant and, hence, that it is indeed appropriate to pool together the two program-specific impact estimates; these statistical tests are not appropriate for making inferences about the true difference in impacts between the two interventions.

The Supplemental Literacy Interventions

The ERO study is a test of supplemental literacy interventions that are designed as full-year courses and targeted to students whose reading skills are two or more years below grade level as they enter high school. Two programs — Reading Apprenticeship Academic Literacy (RAAL), designed by WestEd, and Xtreme Reading, designed by the University of Kansas Center for Research on Learning — were selected for the study from a pool of 17 applicants by a national panel of experts on adolescent literacy. To qualify for the project, the programs were required to focus instruction in the following areas: (1) student motivation and engagement; (2) reading fluency, or the ability to read quickly, accurately, and with appropriate expression; (3) vocabulary, or word knowledge; (4) comprehension, or making meaning from text; (5) phonics and phonemic awareness (for students who could still benefit from instruction in these areas); and (6) writing. The overarching goals of both programs are to help ninth-grade students adopt the strategies and routines used by proficient readers, improve their comprehension skills, and be motivated to read more and to enjoy reading. Both programs are supplemental in that they consist of a yearlong course that replaces a ninth-grade elective class, rather than a core academic class, and in that they are offered in addition to students’ regular English language arts classes.

The primary differences between the two literacy interventions selected for the ERO study lie in their approach to implementation. Implementation of RAAL is guided by the concept of “flexible fidelity” — that is, while the program includes a detailed curriculum, the teachers are trained to adapt their lessons to meet the needs of their students and to supplement program materials with readings that are motivating to their classes. Teachers have flexibility in how they include various aspects of the RAAL curriculum in their day-to-day teaching activities, but they have been trained to do so such that they maintain the overarching spirit, themes, and goals of the program in their instruction.

Implementation of Xtreme Reading is guided by the philosophy that the presentation of instructional material — particularly the order and timing with which the lessons are presented — is of critical import to students’ understanding of the strategies and skills being taught. As such, teachers are trained to deliver course content and materials in a precise, organized, and systematic fashion designed by the developers. Xtreme Reading teachers follow a prescribed implementation plan, following specific day-by-day lesson plans in which activities have allotted segments of time within each class period. Teachers also use responsive instructional practices to adapt and adjust to student needs that arise as they move through the highly structured curriculum.

Overview of the Study

Interventions. Reading Apprenticeship Academic Literacy (RAAL) and Xtreme Reading — supplemental literacy programs designed as full-year courses to replace a ninth-grade elective class. The programs were selected through a competitive applications process based on ratings by an expert panel.

Study sample. Two cohorts of ninth-grade students from 34 high schools and 10 school districts (2,916 students in Cohort 1 and 2,679 students in Cohort 2). Districts and schools were selected by ED’s Office of Vocational and Adult Education through a special Small Learning Communities grant competition. Students were selected based on reading comprehension test scores that were between two and five years below grade level.

Research design. Within each district, high schools were randomly assigned to use either the RAAL program or the Xtreme Reading program during two school years (2005-2006 and 2006-2007). Within each high school, students were randomly assigned to enroll in the ERO class or to remain in a regularly scheduled elective class. A reading comprehension test and a survey were administered to students in the spring of eighth grade or at the start of ninth grade, prior to random assignment, and again at the end of ninth grade. Classroom observations in the first and second semester of the school year were used to measure implementation fidelity.

Outcomes. Reading comprehension and vocabulary test scores, reading behaviors, student attendance in the ERO classes and other literacy support services, implementation fidelity.

The ERO Evaluation

The supplemental literacy programs were implemented in 34 high schools from 10 school districts across the country. The districts were selected through a special grant competition organized by the U.S. Department of Education’s Office of Vocational and Adult Education (OVAE). Experienced, full-time English/language arts or social studies teachers were self-selected and approved by ED, the districts, and the schools to teach the programs for a period of two years.

The ERO evaluation utilizes a two-level random assignment research design. First, within each district, eligible high schools were randomly assigned prior to the first year of program implementation to use one of the two supplemental literacy programs: 17 of the high schools were assigned to use RAAL, and 17 schools were selected to use Xtreme Reading. Each school implemented the same program in two school years: 2005-2006 and 2006-2007. In the second stage of the study design, eligible students within each of the participating high schools and in each year of the study were randomly assigned either to enroll in the ERO class

(the “ERO group”) or to take one of their school’s regularly offered elective classes (the “non-ERO group”).

During the second year of the study, the participating high schools identified 2,679 ninth-grade students with baseline test scores indicating that they were reading two to five years below grade level (an average of 79 students per school). Approximately 57 percent of these students were randomly assigned to enroll in the ERO class, and the remaining students make up the study’s control group and were enrolled in or continued in a regularly scheduled elective class.

Evaluation data were collected with the Group Reading Assessment and Diagnostic Examination (GRADE) reading comprehension and vocabulary tests and a survey.⁵ Both instruments were administered to students at two points in time: a baseline assessment and survey in the spring of eighth grade and a follow-up assessment and survey at the end of ninth grade.⁶ Follow-up test scores are available for 2,171 (81 percent) of the students in the study sample. To learn about the fidelity of program implementation, the study also includes observations of the supplemental literacy classes during the first and second semester of the school year.

Second-Year Implementation

Each ERO teacher (one per school) was responsible for teaching four sections of the ERO class. Each section accommodated between 10 and 15 students. Classes were designed to meet for a minimum of 225 minutes per week and were scheduled as a 45-minute class every day or as a 75- to 90-minute class that met every other day.

- **Of the 34 teachers who participated in the second year of the study, 25 had taught the entire first year of the study, and two had taught a portion of the first year (having replaced a teacher midyear). Seven teachers were new to the ERO programs at the start of the second year.**

During the second year of the project, the developers for each of the ERO programs provided three types of training and technical assistance to both new and returning ERO teachers: a three-day summer training institute in July or August 2006, booster training sessions during the 2006-2007 school year, and three 2-day coaching visits during the 2006-2007 school year. Prior to the summer institute, teachers new to the ERO programs also attended additional

⁵American Guidance Service, *Group Reading Assessment and Diagnostic Evaluation: Teacher’s Scoring and Interpretive Manual, Level H*; and *Technical Manual* (Circle Pines, MN: American Guidance Service, 2001a, 2001b).

⁶In four of the 34 participating schools, baseline testing occurred in the fall of ninth grade rather than the spring of eighth grade.

training sessions at which they were taught the central strategies of the program being implemented in their school.

The study team assessed the overall fidelity with which the ERO programs were implemented in each school during the second year of the project. In the context of this study, “fidelity” refers to the degree to which the observed operation of the ERO program in a given high school was aligned with the intended learning environment and instructional practices that were specified by the model’s developers. The analysis of implementation fidelity in the second year of the study is based on two field research visits to each of the 34 high schools — one during the first semester and one during the second semester of the 2006-2007 school year. The classroom observation protocols used in the site visits provided a structured process for observers to rate the characteristics of the ERO classroom learning environments and the use of ERO instructional strategies by teachers. The instrument included ratings for six characteristics (referred to as “constructs” from here forward) that are common to both programs, as well as ratings for seven program-specific constructs. For each construct, a category rating of 1 (“poorly aligned”), 2 (“moderately aligned”), or 3 (“well aligned”) was given.

The analysis of the classroom observation ratings sought to capture implementation fidelity on two key overarching dimensions of both programs: the classroom learning environment and the teacher’s use of instructional strategies focused on reading comprehension. A composite measure of implementation fidelity was calculated for each of these two dimensions by averaging across the relevant characteristics in the observation protocol. A composite rating of 2.0 or higher indicates that the school’s ERO program was well aligned with the developers’ implementation specifications; a rating of 1.5 to 1.9 means that the program was moderately aligned; and a rating of 1.0 to 1.4 means that it was poorly aligned. Following is a summary of key findings.

- **At the spring site visit, implementation fidelity in 26 of the 34 schools was classified as well aligned on both program dimensions. In seven schools, implementation was classified as moderately aligned with the program model on at least one of the two key program dimensions and as moderately or well aligned on the other dimension. In one school, implementation was deemed to be poorly aligned with the program models.**

The overall implementation of the ERO program in a given school was classified as well aligned if both the classroom environment and the comprehension instruction dimension were rated as being well aligned. According to the protocols used for the classroom observations, teacher behaviors and classroom activities in these schools were consistently rated as being well developed and reflective of the behaviors and activities specified by the developers. At

the fall site visit, the implementation of the ERO programs in 20 of the 34 schools was classified as well aligned on both program dimensions, and, at the spring site visit, 26 schools had attained this benchmark. Because implementation fidelity in the majority of the study schools was deemed to be well aligned to the models, the study team also examined the number of schools whose implementation of the programs was “very well aligned” to developers’ specifications (defined here as a composite score of 2.5 or higher on both program dimensions). At the spring site visit, implementation in 13 schools could be classified as such.

Conversely, a school’s overall implementation fidelity was judged to be poorly aligned with the program model if the composite rating for either the classroom learning environment dimension or the comprehension instruction dimension was rated as poorly aligned. The ERO programs in these schools were not representative of the activities and practices intended by the respective program developers and were found to have encountered serious implementation problems on at least one of the two key program dimensions during the second year of the study.⁷ At the fall site visit, implementation of the ERO programs in three of the 34 schools was classified as poorly aligned with the program models on at least one of the two program dimensions. At the spring site visit, implementation at one school was considered to be poorly aligned with the program models.⁸

- **The number of schools considered to be well aligned with the program developers’ specifications for implementation fidelity was greater in the second year of the study than in the first year (26 schools in the second year, compared with 16 schools in the first year).**

At the spring site visit in the second year of the study, the ERO programs in 33 of the 34 schools reached an overall level of implementation fidelity that was at least moderately aligned to the program models (of these, 26 were considered to be well aligned). This is an improvement over the first year of the study, when 24 of the 34 schools had reached a moderate level of alignment at the spring site visit (of these, 16 schools were deemed to be well aligned). Also, during the spring site visit of the second year, only one school’s implementation of the program was poorly aligned to the developers’ specifications. This is lower than what was found during the first-year spring site visit, when 10 schools were ranked as poorly aligned on at least one of the two key program dimensions.

⁷In particular, poorly aligned implementation for a given dimension means that the classroom observers found that at least half of the classroom characteristics were not aligned with the behaviors and activities specified by the developers and described in the protocols.

⁸In the second year of the study, implementation-fidelity ratings were similar for the 25 schools where the ERO teacher taught two full years of the program and for the nine schools where the ERO teacher had replaced another teacher at some point during the study (an average rating of 2.5 for returning teachers and 2.4 for replacement teachers, out of a maximum of score 3).

Student Enrollment and Attendance in the ERO Classes and Participation in Literacy Support Activities

The study team collected data on the duration of the ERO classes as well as the frequency with which students attended the ERO classes and participated in other classes or tutoring services that aimed to improve their reading and writing skills.

ERO classes in the second year began an average of 2.3 weeks after the start of the school year and operated for an average of nine months. Eighteen schools started the ERO program on the first day of school, and five more schools started within the first two weeks that classes were in session. The remaining eleven started their ERO programs an average of seven weeks after the start of the school year. Among the students randomly assigned to the ERO group, 91 percent enrolled in the ERO classes, and 87 percent were still attending the classes at the end of the school year.

- **Students in the ERO group attended 79 percent of the scheduled ERO classes, and they received an average of 11 hours of ERO instruction per month.**
- **Students who were randomly assigned to the study’s ERO group reported a higher frequency of participation in supplemental literacy services than students who were assigned to the non-ERO group.**

The ERO classes served as the primary source of literacy support services for students in the study sample. Although the largest difference in the use of supplemental literacy supports between the study’s ERO and non-ERO groups occurred in students’ participation in a supplementary school-based literacy class (an average of 75 yearly sessions for ERO students and 17 yearly sessions for non-ERO students), ERO students were also significantly more likely to report working with a tutor in school (an average of 30 yearly sessions, compared with 12 yearly sessions for non-ERO students).

Impact Findings

The GRADE assessment was used to measure students’ reading achievement prior to random assignment (at “baseline”) and then again in the spring at the end of their ninth-grade year (at “follow-up”). The GRADE is a norm-referenced, research-based reading assessment that is used widely to measure performance and track the growth of an individual student and groups of students. Because the two ERO programs focus primarily on helping students use contextual clues to understand the meaning of words, the reading comprehension subtest of the GRADE is the primary measure of reading achievement in this study, while the GRADE vocabulary subtest is a secondary indicator of the programs’ effectiveness. Performance levels and

impacts on both subtests are presented in standard score units; students with a standard score of 100 points are considered to be reading at grade level.⁹

Following is a summary of the study's impact findings.

- **When analyzed jointly, the ERO programs produced an increase of 0.8 standard score point on the GRADE reading comprehension subtests. This corresponds to an effect size of 0.08 standard deviation and is statistically significant. The overall impact of the programs in the second year of implementation is not statistically different from their overall impact in the first year of implementation (0.09 standard deviation).**

The top panel of Table ES.1 shows the impacts on spring follow-up reading comprehension and vocabulary test scores across all 34 participating high schools in the second year of the study. The first row of data in the table shows that, on average, the reading comprehension test scores of students in the ERO group are 0.8 standard score point higher than the scores of students in the non-ERO group, which represents a statistically significant impact (its p-value is less than or equal to 5 percent).¹⁰ Expressed as a proportion of the overall variability of test scores for students in the non-ERO group, this estimated impact represents an effect size of 0.08 (or 8 percent of the standard deviation of the non-ERO group's test scores).

Figure ES.1 places this impact estimate in the context of the actual and expected change in the ERO students' reading comprehension test scores on the GRADE from the beginning of ninth grade to the end of ninth grade. The bottom section of the bar shows that students in the ERO group achieved an average standard score of 84.6 at the start of their ninth-grade year. This corresponds, approximately, to a grade equivalent of 4.9 (the last month of fourth grade) and indicates an average reading level at the 14th percentile for ninth-grade students nationally.

The middle section of the bar shows the estimated growth in test scores experienced by the non-ERO group. At the end of the ninth-grade year, the non-ERO group was estimated to have achieved an average standard score of 89.3, which corresponds to a grade equivalent of 6.0 and an average reading level at the 23rd percentile for ninth-grade students nationally. This

⁹Based on the national norms used to calculate these scores, a standard score of 100 on the GRADE reading comprehension or vocabulary test is average for a representative group of students at the end of their ninth-grade year. The standard deviation of the standard score for both tests is 15.

¹⁰The impact estimates in Table ES.1 are regression-adjusted using ordinary least squares (OLS), controlling for blocking of random assignment by school and for random differences between the ERO and non-ERO groups in their baseline reading comprehension test scores and age at random assignment. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment.

The Enhanced Reading Opportunities Study

Table ES.1

**Impacts on Reading Achievement,
Cohort 2 Follow-Up Respondent Sample**

Outcome	ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>All schools</u>					
Reading comprehension					
Average standard score	90.1	89.3	0.8 *	0.08 *	0.042
<i>Corresponding grade equivalent</i>	6.1	6.0			
<i>Corresponding percentile</i>	25	23			
Reading vocabulary					
Average standard score	93.5	93.5	0.0	0.00	0.986
<i>Corresponding grade equivalent</i>	7.8	7.8			
<i>Corresponding percentile</i>	32	32			
Sample size	1,264	907			
<u>Reading Apprenticeship Academic Literacy schools</u>					
Reading comprehension					
Average standard score	90.2	88.9	1.4 *	0.14 *	0.015
<i>Corresponding grade equivalent</i>	6.1	5.9			
<i>Corresponding percentile</i>	25	23			
Reading vocabulary					
Average standard score	93.4	93.8	-0.4	-0.04	0.428
<i>Corresponding grade equivalent</i>	7.7	7.8			
<i>Corresponding percentile</i>	32	33			
Sample size	645	470			
<u>Xtreme Reading schools</u>					
Reading comprehension					
Average standard score	90.0	89.7	0.2	0.02	0.672
<i>Corresponding grade equivalent</i>	6.1	6.0			
<i>Corresponding percentile</i>	25	24			
Reading vocabulary					
Average standard score	93.5	93.1	0.4	0.04	0.468
<i>Corresponding grade equivalent</i>	7.8	7.7			
<i>Corresponding percentile</i>	32	31			
Sample size	619	437			

(continued)

Table ES.1 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study follow-up GRADE assessment.

NOTES: The follow-up GRADE assessment was administered in the spring of 2007 near the end of students' ninth-grade year.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school and for random differences between the ERO and non-ERO groups in their baseline reading comprehension test scores and age at random assignment. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment.

The national average for standard score values is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE *Teacher's Scoring and Interpretive Manual* (Level H, Grade 9, Spring Testing, Form B). No statistical tests or arithmetic operations were performed on these reference points.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (reading comprehension = 10.035; reading vocabulary = 9.827).

A two-tailed t-test was applied to the impact estimate. The statistical significance is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

growth of 4.7 standard score points for the non-ERO group provides the best indication of what the ERO group would have achieved during their ninth-grade year had they not had the opportunity to attend the ERO classes.

The top section of the bar shows the estimated impact of the ERO programs on reading comprehension test scores. At the end of the ninth-grade year, the ERO group achieved an average standard score of 90.1, which corresponds to a grade equivalent of 6.1 and an average reading level at the 25th percentile for ninth-grade students nationally. This means that the ERO group experienced a growth of 5.5 points in their reading comprehension skills over the course of ninth grade, which is 0.8 point higher than the growth achieved by the non-ERO group. Thus, the impact of the ERO programs (0.8 standard score point) represents a 17 percent improvement over and above the growth that the ERO group would have experienced if they had not had the opportunity to attend the ERO classes (4.7 points).¹¹

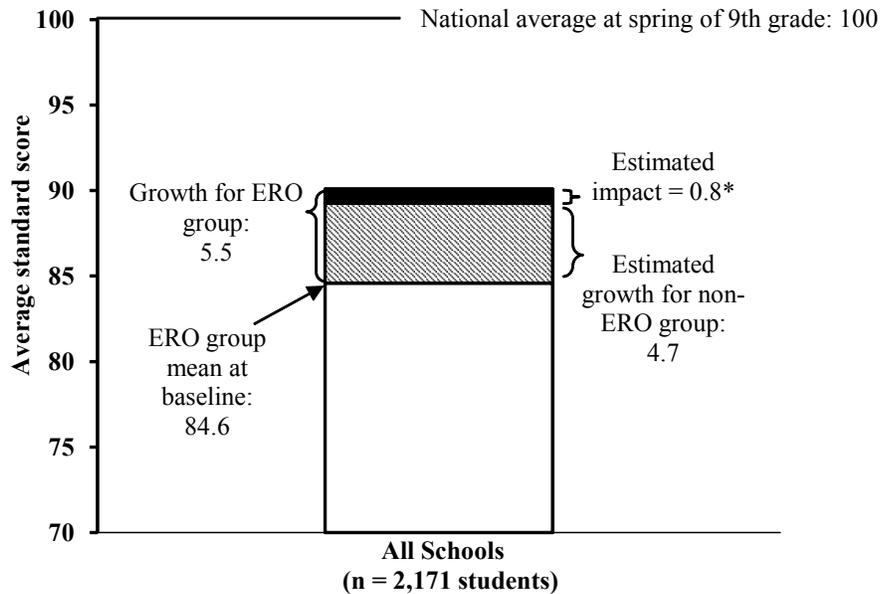
The solid line at the top of Figure ES.1 shows the national average (100 standard score points) for students at the end of ninth grade, in the spring. Students scoring at this level are considered to be reading at grade level. Thus, the ERO group's reading comprehension scores

¹¹The value of 17 percent was calculated by dividing the impact (0.8 standard score point) by the average improvement of the non-ERO group (4.7 standard score points).

The Enhanced Reading Opportunities Study

Figure ES.1

Impacts on Reading Comprehension, Cohort 2 Follow-Up Respondent Sample



SOURCES: MDRC calculations from the Enhanced Reading Opportunities Study baseline and follow-up GRADE assessments.

NOTES: The baseline GRADE assessment was administered in the fall of 2006 at the start of students' ninth-grade year and prior to their random assignment to the ERO and non-ERO groups. The follow-up GRADE assessment was administered in the spring of 2007 near the end of students' ninth-grade year.

The ERO group growth at follow-up is calculated as the difference between the unadjusted ERO group mean at baseline and the unadjusted ERO group mean at follow-up. The impact was estimated using ordinary least squares and adjusted to account for the blocking of random assignment by school and to control for random differences between the ERO and non-ERO groups in baseline reading comprehension test scores and age at random assignment. The expected ERO group growth at follow-up is the difference between the actual ERO group growth and the impact.

A two-tailed t-test was applied to the impact estimate. The statistical significance is indicated (*) when the p-value is less than or equal to 5 percent.

The national average for standard score values is 100, and its standard deviation is 15.

Rounding may cause slight discrepancies in calculating sums and differences.

still lagged nearly 10 points below the national average. In fact, 77 percent of students who participated in the ERO classes scored two or more years below grade level at the end of their ninth-grade year,¹² which means that they would still be eligible for the ERO programs were these programs again made available to them.¹³

- **The RAAL program increased students' reading comprehension test scores by a statistically significant amount (0.14 standard deviation). Although not statistically significant, an impact of 0.2 standard score point on reading comprehension (0.02 standard deviation) was produced by the Xtreme Reading program. The difference in impacts between the two programs is not statistically significant, and thus it cannot be concluded that RAAL had a different effect than Xtreme Reading. Nor is there a statistically significant difference between each program's impact in the second year of implementation and its impact in the first year of implementation.**

The ERO student follow-up survey was administered to students at the same time as the follow-up GRADE assessment and includes additional information on students' reading behaviors and attitudes. Responses to the follow-up survey were used to derive measures for three reading behaviors that are intended to be affected by the ERO programs: the number of times during the prior month that a student read different types of text in school or for homework, the number of times during the prior month that a student read different types of text outside of school, and students' reported use of the reading strategies and techniques that the ERO programs try to teach. The overall impact of the programs on students' reading behaviors is not statistically significant.¹⁴

The Relationship Between Impacts and Second-Year Implementation

This report also includes an exploratory analysis that investigates the relationship between school-level impacts and various aspects of implementation in the second year of the

¹²Forty percent of ninth-graders nationally would be expected to score two years or more below grade level on the GRADE administered in the spring of ninth grade.

¹³Furthermore, 87 percent of the students in the ERO group had reading comprehension scores that were below grade level at the end of ninth grade.

¹⁴The analysis also examines the extent to which impacts on reading comprehension test scores vary across schools. The impact estimates for each school range from a negative impact of 3.7 standard score points to a positive impact of 6.2 standard score points. However, the variation in observed school-level impacts is not statistically significant, indicating that the observed school-to-school variation in impacts may be due to estimation error and may not truly vary across schools.

study. Specifically, this analysis examines whether there are differences in impacts between subgroups of schools defined by teachers' experience with the ERO program (that is, schools whose ERO teacher taught two full years of the program versus schools whose ERO teacher did not teach two full years of the program); overall implementation fidelity during the spring site visit (that is, very well-aligned, well-aligned, moderately aligned, or poorly aligned implementation); and the number of weeks between the start of the school year and ERO program startup (schools that started operating their ERO program within two weeks versus those whose program startup was delayed by two weeks or more). The exploratory analysis also examines whether there are differences in impacts between schools whose implementation of the programs was particularly exemplary (that is, schools that started operating their programs within two weeks *and* whose implementation was very well aligned to the program models) and schools that did not meet these two criteria.¹⁵ *Based on these exploratory analyses, one cannot conclude that the programs were more effective in schools with more experienced ERO teachers, with implementation better aligned with the program models, or with early program startup.* That is, one cannot infer with certainty that these particular implementation characteristics are related to program impacts because the difference in impacts between the groups of schools within each of the three measured categories of implementation — teacher experience teaching the ERO classes, the alignment of the programs as implemented to the program models, and the efficiency of program startup — is not statistically significant. Impacts for the groups of schools with the most promising implementation characterizations are positive and statistically significant (that is, for the 25 schools whose ERO teacher returned in the second year, having taught the entire first year of the program; the 13 schools where the ERO programs were rated as very well aligned to the program models; and the 23 schools where the ERO programs began within the first two weeks of school).¹⁶ Impacts for the related groups of schools with less promising implementation characterizations are smaller and not statistically significant (that is, for the 9 schools whose teachers taught ERO for less than two full years, the 21 schools where there was weaker implementation fidelity, and the 11 schools with program startup that took longer than two weeks). The difference in impacts between the groups of schools within each of the three categories of implementation is not statistically significant.

¹⁵It is important to note that these analyses are exploratory and are not able to establish causal links between these aspects of implementation and variation in program impacts across sites, because other school characteristics and implementation factors may confound the association between school-level impacts and the implementation factors included in the exploratory analysis.

¹⁶The impacts on reading comprehension test scores for each of these three groups of schools are as follows: in the 25 schools whose ERO teacher had returned having taught all of the first year of the program, the effect size is 0.09 standard deviation (p-value = 0.050); in the 13 schools where implementation was rated as very well aligned to the program models, the effect size is 0.13 standard deviation (p-value = 0.047); and in the 23 schools where the programs began within the first two weeks of school, the effect size is 0.10 standard deviation (p-value = 0.048).

Next Steps for the ERO Study

The ultimate goal of the two ERO programs is to improve students' academic performance during high school and to keep them on course toward graduation. With this in mind, the final report from the evaluation — scheduled for 2009 — will examine the impact of the programs on the achievement and attainment outcomes of both cohorts of students as they progress through high school. The outcomes examined in the report will include students' performance in core academic classes, their performance on the high-stakes tests required by their states, their grade-to-grade promotion rates, and whether they are on track to graduate from high school.

References

- American Guidance Service. 2001a. *Group Reading Assessment and Diagnostic Evaluation: Teacher's Scoring and Interpretive Manual, Level H*. Circle Pines, MN: American Guidance Service.
- American Guidance Service. 2001b. *Group Reading Assessment and Diagnostic Evaluation: Technical Manual*. Circle Pines, MN: American Guidance Service.
- Kemple, James J., William Corrin, Elizabeth Nelson, Terry Salinger, Suzannah Herrmann, and Kathryn Drummond. 2008. *The Enhanced Reading Opportunities Study: Early Impacts and Implementation Findings* (NCEE 2008-4015). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.