

The Enhanced Reading Opportunities Study

Early Impact and Implementation Findings

The Enhanced Reading Opportunities Study

Early Impact and Implementation Findings

JANUARY 2008

James J. Kemple
William Corrin
Elizabeth Nelson
MDRC

Terry Salinger
Suzannah Herrmann
Kathryn Drummond
American Institutes for Research

Paul Strasberg, Project Officer
Institute of Education Sciences

NCEE 2008-4015
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

January 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: Kemple, J., Corrin, W., Nelson, E., Salinger, T., Herrmann, S., and Drummond, K. (2008). *The Enhanced Reading Opportunities Study: Early Impact and Implementation Findings* (NCEE 2008-4015). 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.

Contents

List of Exhibits	v
Acknowledgments	ix
Disclosure of Potential Conflicts of Interest	x
Executive Summary	xi
Chapter	
1 Introduction	1
Striving Adolescent Readers: The Nature and Consequences of the Problem	3
Key Elements of a Response and the Role of Supplemental Literacy Programs	5
Overview of the ERO Study	8
Overview of This Report	11
2 Study Sample and Design	13
School Sample	13
Student Sample	15
Data Sources and Measures	20
Follow-Up Data Collection and Response Rates	27
Analytic Methods and Procedures	31
3 Implementing the Supplemental Literacy Programs	39
Characteristics of the Supplemental Literacy Programs: Reading Apprenticeship	
Academic Literacy and Xtreme Reading	40
The ERO Teachers and Their Preparation for the ERO Programs	46
Implementation Fidelity	50
Summary and First-Year Implementation Challenges	57
4 Student Attendance in the ERO Classes and Participation in Literacy Support Activities	
Student Enrollment and Attendance in the ERO Classes	61
Student Participation in Literacy Support Activities	62
5 Early Impacts on Student Reading Achievement and Reading Behaviors	71
Early Impacts on Reading Achievement	73
Early Impacts on Students' Reading Behaviors	77
Early Impacts for Subgroups of Students	79
The Relationship between Early Impacts and First-Year Implementation Issues	84
Conclusion	91

Appendixes

A	ERO Student Follow-Up Survey Measures	93
B	Follow-Up Test and Survey Response Analysis	115
C	Statistical Power and Minimum Detectable Effect Size	135
D	ERO Implementation Fidelity	141
E	Technical Notes for Early Impact Findings	173
F	Early Impact Estimates Weighted for Nonresponse	185
G	Early Impacts on Supplementary Measures of Reading Achievement and Behaviors	191
H	Early Impacts for Student Subgroups	197
I	The Relationship Between Early Impacts and First-Year Implementation Issues	213
References		231

List of Exhibits

Table

ES.1	Impacts on Reading Achievement, Cohort 1 Follow-Up Respondent Sample	xviii
ES.2	Impacts on Reading Achievement, Cohort 1 Follow-Up Respondent Sample, by First-Year Implementation Issues	xxii
2.1	Characteristics of ERO Schools and Average Schools in the United States (2004-2005)	16
2.2	Characteristics of Students in Cohort 1 Full Study Sample	21
2.3	Response Rates of Students in Cohort 1 Full Study Sample	28
2.4	Characteristics of Students in Cohort 1 Follow-Up Respondent Sample	30
2.5	Characteristics of Students in Cohort 1 Follow-Up Respondent Sample, Reading Apprenticeship Schools	32
2.6	Characteristics of Students in Cohort 1 Follow-Up Respondent Sample, Xtreme Reading Schools	34
3.1	Key Components of the ERO Programs	42
3.2	Background Characteristics of ERO Teachers	47
3.3	Training and Technical Assistance Provided During the 2005-2006 School Year, by ERO Program	48
3.4	Dimensions and Component Constructs of Implementation Fidelity, by ERO Program	52
3.5	Number of ERO Classrooms Well, Moderately, or Poorly Aligned to Program Models on Each Implementation Dimension, by ERO Program	54
4.1	Attendance in ERO Classes, Follow-Up Respondent Sample in the ERO Group	64
4.2	Comparison of ERO and Non-ERO Student Schedules	67
4.3	Participation in Supplemental Literacy Support Activities, Cohort 1 Follow-Up Respondent Sample	70
5.1	Impacts on Reading Achievement, Cohort 1 Follow-Up Respondent Sample	74
5.2	Impacts on Reading Achievement, Cohort 1 Follow-Up Respondent Sample, by Program	78
5.3	Impacts on Reading Behaviors, Cohort 1 Follow-Up Respondent Sample	80
5.4	Impacts on Reading Behaviors, Cohort 1 Follow-Up Respondent Sample, by Program	81
5.5	Impact Effect Sizes for Student Subgroups	83

Table

5.6	Impact Effect Sizes, by First-Year Implementation Issues	88
A.1	Intensity Values for Supplemental Literacy Support Measures	96
B.1	Response Rates of Students in Cohort 1 Full Study Sample	119
B.2	Characteristics of Students in Cohort 1: Differences Between Respondents and Nonrespondents	120
B.3	Characteristics of Students in Cohort 1: Differences Between Respondents and Nonrespondents, Reading Apprenticeship Schools	122
B.4	Characteristics of Students in Cohort 1: Differences Between Respondents and Nonrespondents, Xtreme Reading Schools	124
B.5	Regression Coefficients for the Probability of Being in the Respondent Sample, Full Study Sample	126
B.6	Regression Coefficients for the Probability of Being in the Treatment Group, Respondent Sample	130
C.1	Sample Sizes, by Site and Student Subgroup Configuration, for Full Sample and 80 Percent Subsample	140
C.2	Minimum Detectable Effect Sizes, by Site and Student Subgroup Configuration, for Full Sample and 80 Percent Subsample	140
D.1	Number of ERO Classrooms Well, Moderately, or Poorly Aligned to Program Models on Each Implementation Dimension, by ERO Program	150
E.1	Impacts on Reading Achievement, Cohort 1 Follow-Up Respondent Sample	175
E.2	Impacts on Reading Behaviors, Cohort 1 Follow-Up Respondent Sample	177
E.3	Impacts on Reading Behaviors Composite Index, for the Full Study Sample and Subgroups	183
F.1	Impacts on Reading Achievement Weighted by School Response Rate, Cohort 1 Follow-Up Respondent Sample	187
F.2	Impacts on Reading Behaviors Weighted by School Response Rate, Cohort 1 Follow-Up Respondent Sample	189
G.1	Impacts on Attitudes and Perceptions of Reading and School, Cohort 1 Follow-Up Respondent Sample	194
G.2	Impacts on Percentage of Students No Longer Eligible for Program, Cohort 1 Follow-Up Respondent Sample	196
H.1	Impacts on Reading Achievement, Cohort 1 Follow-Up Respondent Sample, by Baseline Reading Comprehension Performance	200
H.2	Impacts on Reading Behaviors, Cohort 1 Follow-Up Respondent Sample, by Baseline Reading Comprehension Performance	202

Table

H.3	Impacts on Reading Achievement, Cohort 1 Follow-Up Respondent Sample, by Whether Students Were Overage for Grade	204
H.4	Impacts on Reading Behaviors, Cohort 1 Follow-Up Respondent Sample, by Whether Students Were Overage for Grade	206
H.5	Impacts on Reading Achievement, Cohort 1 Follow-Up Respondent Sample, by Language Spoken at Home	208
H.6	Impacts on Reading Behaviors, Cohort 1 Follow-Up Respondent Sample, by Language Spoken at Home	210
I.1	Fixed-Effect Impact Estimates on Reading Comprehension, by School	216
I.2	Impacts on Reading Achievement, Cohort 1 Follow-Up Respondent Sample, by Program Implementation Fidelity	218
I.3	Impacts on Reading Behaviors, Cohort 1 Follow-Up Respondent Sample, by Program Implementation Fidelity	220
I.4	Impacts on Reading Achievement, Cohort 1 Follow-Up Respondent Sample, by Program Duration	222
I.5	Impacts on Reading Behaviors, Cohort 1 Follow-Up Respondent Sample, by Program Duration	224
I.6	Impacts on Reading Achievement, Cohort 1 Follow-Up Respondent Sample, by First-Year Implementation Issues	226
I.7	Impacts on Reading Behaviors, Cohort 1 Follow-Up Respondent Sample, by First-Year Implementation Issues	228

Figure

ES.1	Impacts on Reading Comprehension, Cohort 1 Follow-Up Respondent Sample	xx
2.1	Construction of the Impact Sample from the Eligibility Pool	19
5.1	Impacts on Reading Comprehension, Cohort 1 Follow-Up Respondent Sample	75
5.2	Fixed-Effect Impact Estimates on Reading Comprehension, by School	86

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, allow access to classrooms, and 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, Courtney Zmach, Rebecca Holland-Coviello, Christopher Doyle, and Andrea Olinger coordinated the classroom observations and supported the observation staff. Nancy Lang and Courtney Tanenbaum processed and managed the interview and observation data. These AIR staff also conducted classroom observations and phone interviews.

At MDRC, Susan Sepanik and Edmond Wong assisted with data collection and provided programming and analysis support. Corinne Herlihy and Kristin Porter served as school district coordinators. Gordon Berlin, Howard Bloom, Fred Doolittle, Corinne Herlihy, Janet Quint, and Pei Zhu provided substantive expertise through their thoughtful comments on, and reviews of, this report. Mario Flecha and Vivan Mateo assisted with report production. Robert Weber and John Hutchins 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 early 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 first of three — focuses on the first of two cohorts of ninth-grade students who will 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 first year of the study and provides an assessment of the overall fidelity with which the participating schools adhered to the program design specified by the developers. 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.** This impact estimate is statistically significant. Despite the improvement in reading comprehension, 76 percent of the students who enrolled in the ERO classes were still reading at two or more years below grade level at the end of ninth grade.
- **Although they are not statistically significant, the magnitudes of the impact estimates for each literacy intervention are the same as those for the full study sample.**
- **Impacts on reading comprehension are larger for the 15 schools where (1) the ERO programs began within six weeks of the start of the school year and (2) implementation was classified as moderately or well aligned with the program model, compared with impacts for the 19 schools where at least one of these conditions was not met.** The difference in impacts on reading comprehension between these two groups of schools is statistically significant. It is important to note, however, that these two factors

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

did not necessarily cause the differences in impacts and that other factors may be also associated with differences in estimated impacts across schools.

The next report from the study — scheduled for 2008 — will provide findings for a second year of program implementation and a second cohort of ninth-grade students who are enrolled in the ERO classes. 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 ERO programs for both cohorts of students on their performance in core academic classes, their grade-to-grade promotion rates, and their performance on high-stakes tests required by their states.

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, 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 year-long 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 Reading Apprenticeship Academic Literacy 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 Reading Apprenticeship curriculum in their day-to-day teaching activities, but 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.

Study Overview

Interventions: Reading Apprenticeship Academic Literacy 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: 2,916 ninth-grade students from 34 high schools and 10 school districts. 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 Reading Apprenticeship Academic Literacy program or the Xtreme Reading program. 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 at the start of ninth grade prior to random assignment and at the end of ninth grade. Classroom observations in the 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 are being 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 to use one of the two supplemental literacy programs: 17 of the high schools were assigned to use Reading Apprenticeship Academic Literacy, and 17 schools were selected to use Xtreme Reading.

The second feature of the study design involves the random assignment of eligible and appropriate students within each of the participating high schools. During the first year of the study, the participating high schools identified an average of 85 ninth-grade students who were reading at least two years below grade level. Approximately 55 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. The first cohort of the study sample includes 2,916 ninth-grade students with baseline test scores indicating that they were reading between the fourth- and seventh-grade levels.

Evaluation data were collected with the Group Reading Assessment and Diagnostic Examination (GRADE) reading comprehension and vocabulary tests and a survey.² Both instruments were administered to students at two points during the ninth-grade year: a baseline assessment and survey at the start of ninth grade and a follow-up assessment and survey at the end of ninth grade. Follow-up test scores and surveys are available for 2,413 (83 percent) of the students in the study sample. To learn about the fidelity of program implementation, the study also included observations of the supplemental literacy classes during the second semester of the school year.

First-Year Implementation

During the first year of the project, the developers for each of the ERO programs provided three types of training and technical assistance to one teacher from each of the 34 participating schools who volunteered to teach the ERO classes: a five-day summer training institute in August 2005, booster training sessions during the 2005-2006 school year, and a minimum of two 1-day coaching visits during the 2005-2006 school year.

Each ERO teacher 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. The ERO classes began an average of six weeks after the start of the 2005-2006 school year, with the earliest programs starting three weeks into the

²American 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).

school year and the latest programs starting 10 weeks into the school year. The late start was due to the fact that the process for identifying eligible students for the program could not begin until the start of the school year and required extensive effort on the part of school staff and the study team to help complete the baseline data collection process and gain consent from students and their parents.

The study team assessed the overall fidelity with which the ERO programs were implemented in each school during the first 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 environments and instructional practices that were specified by the model developers. Measures of implementation fidelity were developed from 140 to 180 minutes of observation of each ERO classroom conducted in the second semester of the school year. Composite fidelity scores were calculated from numeric ratings (ranging from one to three) of classroom activities related to two overarching program dimensions: classroom learning environment and comprehension instruction. The implementation fidelity for each dimension was classified as *well aligned*, *moderately aligned*, or *poorly aligned*, based on the composite scores. Following is a summary of key findings.

- **The implementation of the ERO programs in 24 of the 34 schools was classified as well aligned or moderately aligned with their program models on both the classroom learning environment and the comprehension instruction dimensions. This included 11 of the schools using Reading Apprenticeship Academic Literacy and 13 of the schools using Xtreme Reading.**

The implementation of the ERO programs in 16 of the 34 schools was classified as well aligned on both program dimensions. Because the classroom learning environments and comprehension instruction activities were designed to be interdependent and mutually reinforcing, the implementation of ERO program in a given school was classified as well aligned only if both of these dimensions were rated as well aligned. According to the protocols used for the classroom observations, teacher behaviors and classroom activities in these schools were rated consistently as being well developed and reflective of the behaviors and activities specified by the developers.

The implementation of the ERO programs in eight of the 34 schools was classified as moderately aligned with the program model on at least one of the two key program dimensions and moderately or well aligned on the other dimension. In six of these schools, the classroom learning environment was classified as well aligned with the program model while the comprehension instruction was classified as moderately aligned. In the remaining two schools, both the

classroom learning environment and the comprehension instruction were rated as being moderately aligned with their program models.

- **The implementation of the ERO programs in 10 of the 34 schools was classified as poorly aligned with the program models on at least one of the two overarching program dimensions. This includes six of the schools using Reading Apprenticeship Academic Literacy and four of the schools using Xtreme Reading.**

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 inadequate. Poorly aligned implementation for a given dimension meant 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. The ERO programs in these schools were the least 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 first year of the study.

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

The study team collected data on 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.

The ERO classes began an average of six weeks after the start of the school year and operated for an average of seven and a half months of the nine-month school year. More than 95 percent of the students randomly assigned to the ERO group enrolled in the ERO classes, and 91 percent were still attending the classes at the end of the school year.

- **Students in the ERO group attended 83 percent of the scheduled ERO classes, and they received an average of just over 11 hours of ERO instruction per month. Attendance rates were similar for schools using Reading Apprenticeship Academic Literacy (82 percent) and for those using Xtreme Reading (84 percent).**
- **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 use of supplemental literacy supports between the study's ERO and non-ERO groups occurred in students' attendance in school-based literacy class, ERO students were also more likely to report attending a literacy class outside school and working with a tutor in and outside school. According to the student survey, students in the ERO group reported that they attended an average of 52 more school-based class sessions during the year that focused on reading or writing, compared with students in the non-ERO group. Depending on a school's scheduling structure, classes meet between 90 and 180 times per year. Students in the ERO group were also more likely to report attending these types of classes outside school (an average of 3 more sessions reported during the year, compared with the non-ERO group). Finally, students in the ERO group were more likely to report working on their reading and writing with a tutor (an average of 17 more sessions for the year, compared with the non-ERO group). Each of these differences is statistically significant.

Early Impact Findings

The primary measure of reading achievement for the ERO study is students' scores on the reading comprehension assessment subtest of the GRADE. A secondary measure of students' reading achievement is their scores on the GRADE vocabulary assessment. Following is a summary of the study's early impact findings.

- **When analyzed jointly, the ERO programs produced an increase of 0.9 standard score point on the GRADE reading comprehension subtests. This corresponds to an effect size of 0.09 standard deviation and is statistically significant.**

The top panel of Table ES.1 shows the impacts on reading comprehension and vocabulary test scores across all 34 participating high schools. The first row in the table shows that, overall, the ERO programs improved reading comprehension test scores by 0.9 standard score point and that this impact is statistically significant (its p-value is less than or equal to 5 percent). Expressed as a proportion of the overall variability of test scores for students in the non-ERO group, this represents an effect size of 0.09 (or 9 percent of the standard deviation on 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 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 85.9 at the start of their ninth-grade year. This corresponds, approximately, to a grade equivalent of 5.1 (the first month of fifth grade) and indicates an average reading level at the 16th percentile for ninth-grade students nationally. The middle

The Enhanced Reading Opportunities Study

Table ES.1

**Impacts on Reading Achievement,
Cohort 1 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	89.2	0.9 *	0.09 *	0.019
<i>Corresponding grade equivalent</i>	6.1	5.9			
<i>Corresponding percentile</i>	25	23			
Reading vocabulary					
Average standard score	93.4	93.2	0.3	0.03	0.472
<i>Corresponding grade equivalent</i>	7.7	7.7			
<i>Corresponding percentile</i>	32	31			
Sample size	1,408	1,005			
<u>Reading Apprenticeship schools</u>					
Reading comprehension					
Average standard score	89.8	88.9	0.9	0.09	0.097
<i>Corresponding grade equivalent</i>	6.1	5.9			
<i>Corresponding percentile</i>	24	23			
Reading vocabulary					
Average standard score	93.2	92.8	0.5	0.05	0.393
<i>Corresponding grade equivalent</i>	7.7	7.7			
<i>Corresponding percentile</i>	31	31			
Sample size	686	454			
<u>Xtreme Reading schools</u>					
Reading comprehension					
Average standard score	90.5	89.6	0.9	0.09	0.090
<i>Corresponding grade equivalent</i>	6.2	6.0			
<i>Corresponding percentile</i>	25	24			
Reading vocabulary					
Average standard score	93.6	93.5	0.1	0.01	0.846
<i>Corresponding grade equivalent</i>	7.8	7.8			
<i>Corresponding percentile</i>	32	32			
Sample size	722	551			

NOTE: The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

section of the bar shows the estimated growth in test scores experienced by the non-ERO group. This growth of 3.4 standard score points 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. At the end of the ninth-grade year, therefore, the non-ERO group was estimated to have achieved an average standard score of 89.2, which corresponds to a grade equivalent of 5.9 and an average reading level at the 23rd percentile for ninth-grade students nationally.

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 was estimated to have 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. Thus, the impact of the ERO programs represents a 26 percent improvement over and above what the ERO group would have achieved if they had not had the opportunity to attend the ERO classes.³

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. Despite the program impact, therefore, the ERO group's reading comprehension scores still lagged nearly 10 points below the national average. In fact, almost 90 percent of the students in the ERO group had reading comprehension scores that were below grade level at the end of ninth grade. Hence, 76 percent of students who participated in the ERO classes would still be eligible for the programs because they had scored more than two years below grade level at the end of their ninth-grade year.

- **Although neither program-specific impact is statistically significant, estimated impacts for schools using the Reading Apprenticeship Academic Literacy program and for schools using Xtreme Reading are 0.9 standard score point.**

Table ES.1 shows that the impacts on reading comprehension for both Reading Apprenticeship and Xtreme Reading are of similar magnitude to that found for the full sample of schools in the study. Neither of these estimates is statistically significant, however.

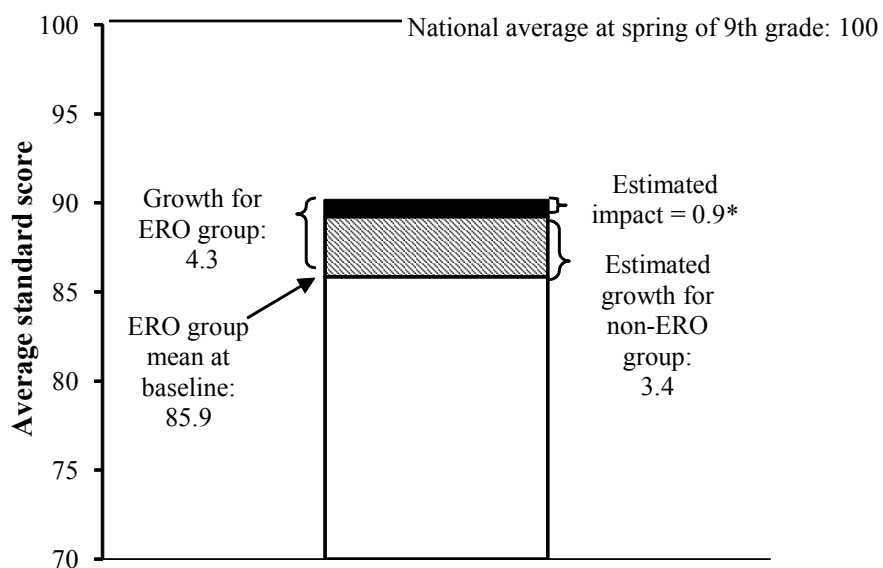
The ERO Student Follow-up Survey included questions about students' reading behavior. The impact analysis focused on three measures that were developed from these questions: the amount of reading students do for school, the amount of reading students do for non-school purposes, and students' use of reflective reading strategies. While the ERO programs produced some changes in these reading behaviors (both positive and negative), none of the estimated impacts is statistically significant.

³This was calculated by dividing the impact (0.9 standard score point) by the average improvement of the non-ERO group (3.4 standard score points).

The Enhanced Reading Opportunities Study

Figure ES.1

Impacts on Reading Comprehension, Cohort 1 Follow-Up Respondent Sample



NOTE: The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

First-Year Implementation Challenges and Early Impacts

The first-year start-up experiences in 19 of the 34 participating high schools were particularly problematic either because of poorly aligned implementation fidelity or because of especially long delays in enrolling students in their ERO classes. Of these, seven high schools experienced poorly aligned implementation, even though they were able to begin the classes within six weeks of the start of the school year, and nine high schools experienced a start-up delay of more than six weeks, even though the implementation of their ERO programs ended up being classified as at least moderately aligned with their program models. The remaining three high schools experienced both poorly aligned implementation and a start-up delay of more than six weeks. The presence of these implementation challenges in 19 of the high schools raises questions about whether the ERO programs had stronger impacts for the 15 high schools that were able to begin classes within six weeks of the start of the school year *and* where implementation was classified as moderately or well aligned with the program models.

Table ES.2 shows the impacts on reading test scores for the two groups of schools defined by their first-year start-up experiences. The top panel of the table shows the impacts for the 15 schools that operated their programs for at least seven and a half months and reached a limited or adequate level of implementation fidelity on both the classroom learning environment and the comprehension instruction dimensions. These ERO programs produced positive and statistically significant impacts on reading comprehension test scores. The schools with a stronger start-up produced an increase of 1.8 standard score points in reading comprehension. This is equivalent to an effect size of 0.17 standard deviation and is statistically significant.

The bottom panel of Table ES.2 presents estimated impacts on reading comprehension test scores for ERO programs in schools where implementation fidelity was found to be inadequate or where the programs operated for seven and a half months or less in the first year. The difference is not statistically significant. The difference between the impact for the stronger start-up schools and the weak start-up schools is 1.6 standard score points and an effect size of 0.16. This difference in impacts is statistically significant, indicating that there is a systematic difference in impacts across these two groups of schools.

It is important to note that the analyses just discussed are exploratory and are not able to establish causal links between these early implementation challenges and variation in program impacts across the sites. There are other school characteristics and implementation factors that may also be associated with variation in estimated impacts. As an exploratory analysis, it is also not appropriate to extrapolate from these findings to determine the impact of the ERO programs in the second year of the project.

Next Steps for the ERO Study

The early impact findings discussed in this report do not represent conclusive evidence about the efficacy or effectiveness of the supplemental literacy interventions being tested. The next report from the ERO study will provide evidence on the impact of the supplemental literacy programs during the second year of implementation. A critical goal of the second year of the implementation has been to build on the experiences of the ERO teachers and the program developers to address the start-up challenges that arose in the first year. Twenty-seven of the 34 teachers who taught the ERO classes in the first year of the study returned for the second year. These teachers participated in a second summer training institute and continued to learn more about how to use the instructional strategies that lie at the heart of the two interventions. The seven new teachers participated in extensive training to help them begin teaching the class with as much fidelity to the model's specifications as possible and have received coaching throughout the year. A second cohort of ninth-grade students entered the study sample in the 2006-2007 school year. Most of the students in the ERO group from this cohort began their enrollment in the ERO classes at or near the start of the school year.

The Enhanced Reading Opportunities Study

Table ES.2

**Impacts on Reading Achievement,
Cohort 1 Follow-Up Respondent Sample,
by First-Year Implementation Issues**

Outcome	Non-ERO ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Moderately or well-aligned implementation and longer duration</u>					
Reading comprehension					
Average standard score	90.7	89.0	1.8 *	0.17 *	0.002
<i>Corresponding grade equivalent</i>	6.2	5.9			
<i>Corresponding percentile</i>	26	23			
Reading vocabulary					
Average standard score	93.6	93.5	0.1	0.01	0.848
<i>Corresponding grade equivalent</i>	7.8	7.7			
<i>Corresponding percentile</i>	32	32			
Sample size	656	488			
<u>Poorly aligned implementation or shorter duration</u>					
Reading comprehension					
Average standard score	89.6	89.5	0.1	0.01	0.811
<i>Corresponding grade equivalent</i>	6.0	6.0			
<i>Corresponding percentile</i>	24	24			
Reading vocabulary					
Average standard score	93.3	92.9	0.4	0.04	0.412
<i>Corresponding grade equivalent</i>	7.7	7.7			
<i>Corresponding percentile</i>	32	31			
Sample size	752	517			
<u>Differences in impacts</u>					
Difference in Impacts Between Subgroups			Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
Reading comprehension standard score			1.6 *	0.16 *	0.035
Reading vocabulary standard score			-0.3	-0.03	0.667

NOTE: The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

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 will examine the impact of the programs on student performance in their core academic classes, their grade-to-grade promotion rates, and their performance on high-stakes tests required by their states. The final report will present impacts on these outcomes through the eleventh grade for students in the study's first cohort and through the tenth grade for students in the second cohort.

Chapter 1

Introduction

According to the National Assessment of Educational Progress (NAEP), just over 70 percent of students nationally arrive in high school with reading skills that are below “proficient” — defined as demonstrating competency over challenging subject matter. Nearly half of these students do not exhibit even partial mastery of knowledge and skills that are fundamental to proficient work at grade level.¹ These limitations in literacy skills are a major source of course failure, high school dropout, and poor performance in postsecondary education.² While research is beginning to emerge about the special needs of striving adolescent readers, very little is known about effective interventions aimed at addressing these needs.³

To help fill this gap and to provide evidence-based guidance to practitioners, the U.S. Department of Education initiated the Enhanced Reading Opportunities (ERO) study — a demonstration and rigorous evaluation of supplemental literacy programs targeted to ninth-grade students with limited literacy skills.⁴ The demonstration involves 34 high schools from 10 school districts that are implementing one of two supplemental literacy programs: Reading Apprenticeship Academic Literacy, designed by WestEd, or Xtreme Reading, designed by the Kansas University Center for Research on Literacy. These programs were selected from a pool of 17 applicants for this project by a national panel of experts on adolescent literacy. The programs are supplemental in that they consist of a year-long course that replaces a ninth-grade elective class rather than a core academic class. They aim to help striving adolescent readers develop the strategies and routines used by proficient readers and to motivate them to read more and to apply these strategies to a wide range of texts.

The evaluation is assessing the impact of the two supplemental literacy programs on students’ reading comprehension skills and on their general performance in high school, including achievement on standardized tests, course completion, and progress toward graduation. MDRC — a nonprofit, nonpartisan social policy research organization — is conducting the evaluation in partnership with the American Institutes for Research (AIR) and Survey Research Management (SRM).

¹Lutkus, Rampey, and Donahue (2006) provide an analysis of NAEP reading results for urban school districts in the context of the national NAEP performance trends.

²Carnevale (2001); Kamil (2003); Snow and Biancarosa (2003).

³Biancarosa and Snow (2004).

⁴The ERO study is known more formally as “An Evaluation of the Impact of Supplemental Literacy Interventions in Freshman Academies.”

The evaluation is based on a two-level random assignment research design. In the first stage, 34 participating high schools were randomly assigned to use one of the two supplemental literacy programs. In the second stage, more than 2,900 eligible students from these high schools (students with reading test scores between two and five years below grade level) were randomly assigned either to participate in one of the literacy programs or to continue in a regular elective class.

Evaluation data were collected with a standardized reading comprehension test and a survey that were administered to students at two points during the ninth-grade year: (1) a baseline assessment and survey at the start of ninth grade and (2) a follow-up assessment and survey at the end of ninth grade. The study also includes observations of the supplemental literacy classes and interviews with teachers and administrators in each of the high schools, to learn about the fidelity of program implementation.

This report presents early findings from the ERO study, based on the first year that the supplemental literacy programs were in operation. It focuses on the first of two cohorts of ninth-grade students from each of the participating high schools. The report assesses the impact that the two supplemental literacy programs had on these students' reading comprehension skills through the end of their ninth-grade year. The report also presents impacts on selected reading behaviors, as a secondary indicator of the programs' potential effect on the initial cohort of students. The report provides an assessment of the fidelity with which the programs were implemented and discusses factors that influenced the capacity of the schools and teachers to operate them as intended over the course of the study's first year.

The early findings presented in this report should be seen as preliminary because of the implementation challenges that arose from the rushed start of the project and that are often typical of the initial phases of complex demonstrations. Also, while the end of ninth grade and the end of students' exposure to the literacy programs is a useful point at which to assess impacts on reading comprehension skills, the evaluation does not yet include information on students' longer-term performance in high school. This means that it is too early to draw definitive conclusions about the potential of these literacy interventions to improve the performance of striving adolescent readers.

In anticipation of these challenges, the U.S. Department of Education extended the demonstration and evaluation to include a second cohort of ninth-grade students who would be exposed to the programs during their second year of operation. Two subsequent reports from the ERO study will provide stronger evidence about program impacts and implementation. The second report will focus on the second year of implementation and on the second cohort of ninth-grade students to enter the study sample. In the second year of the study, most of the schools did not experience the start-up delay that they encountered in the first year. Thus, in

most of the participating schools, findings for the second cohort of students will reflect their exposure to a full year of program operation and to teachers who were more experienced in implementing the programs. The third report will focus on the longer-term impacts on students' academic achievement in tenth and eleventh grades, including their performance on high-stakes state tests and their progress toward graduation.

The remainder of this chapter describes the nature and consequences of the low literacy levels with which many students enter high school — a key motivation for the ERO study. It also provides a more detailed description of the ERO demonstration and of the research design being used to assess the impact of the two supplemental literacy programs selected for the project.

Striving Adolescent Readers: The Nature and Consequences of the Problem

The ERO study emerged from the growing recognition of the role that limited literacy skills play in restricting student success throughout high school and, particularly, during the tenuous transition from eighth to ninth grade. Some view large, comprehensive high schools as impersonal, bureaucratic, anonymous, and unable to respond effectively to the diverse needs of adolescents.⁵ Such schools can be especially inhospitable to ninth-graders — particularly to students with weak academic preparation, especially in literacy — and can exacerbate feelings of low self-efficacy and social marginalization.⁶ Further, as students progress through the primary grades to the middle grades and then to high school, they read increasingly complex textbooks, supplementary materials, and electronic text. In particular, the reading requirements of ninth grade represent a new and giant leap for entering freshmen, who face an increase in the amount of reading that is required in their courses, textbooks that are thicker and more intimidating than in previous grades, and a vocabulary load in content-area instruction that can be overwhelming. Struggling readers — who may harbor real interest in their academic subjects but lack confidence in their ability to improve their reading — may feel uncomfortable in school, may increasingly avoid challenging reading materials, and may try to avoid situations in which their poor reading skills will be exposed.⁷

Recent research indicates that struggling adolescent readers grapple with a constellation of reading difficulties that range from severe problems with basic literacy skills to troubles gaining a nuanced understanding of text. According to a report issued by the Southwest Educational

⁵National Association of Secondary School Principals (1996); Darling-Hammond, Aness, and Ort (2002); Sizer (1984); Harvey and Housman (2004).

⁶Legters and Kerr (2001); Lee, Bryk, and Smith (1993); Shanahan (2004).

⁷Guthrie (2002); Guthrie and Alvermann (1999); Wigfield (2004).

Development Laboratory, struggling adolescent readers generally demonstrate the following characteristics:⁸

1. Their reading is often slow and lacking in fluency, often because they struggle with decoding.
2. Their comprehension skills are weak, often because of limited background knowledge, difficulty making inferences, limited vocabulary, and limited self-regulation strategies.
3. They lack motivation to persist in reading.

In their report *Reading Next — A Vision for Action and Research in Middle and High School Literacy*, Biancarosa and Snow indicate that about “70 percent of older readers require some form of remediation.”⁹ However, these students’ problem is less often with knowing how to read words on a page and rather more often with understanding what they read; that is, they have difficulties with comprehension.¹⁰ Their struggles with comprehension can stem from lack of fluency (they cannot read quickly enough to facilitate comprehension) or from a lack of strategies for how to make sense of what they read or even from a lack of experience employing such strategies across a variety of types of texts in different situations. The goal for these readers is to advance from basic literacy skills to mastering the reading comprehension skills necessary for success in secondary school and beyond. That is, although some adolescent readers may still need support with basic reading skills — decoding, phonics, phonemic awareness, and so on — the majority need additional support and instruction to become expert readers who can move through complex passages containing advanced vocabulary — with fluency and the ability to derive the intended meaning.¹¹

Most high schools provide no formal instructional supports for literacy development, and most English/language arts and social studies teachers do not see literacy development as within their purview. Researchers have noted some common attitudes toward and assumptions about literacy instruction in high schools that may account for this gap. Most significantly, high school teachers view literacy skills as functional tools to be employed in the service of content-area learning.¹² Roe, Stoodt, and Burns suggest that secondary school instructional planning also reflects the belief that teaching reading is the domain of elementary schools, that teaching reading in the content areas is separate from teaching subject matter, that teaching reading in

⁸Peterson et al. (2000).

⁹Biancarosa and Snow (2004) focus on students in grades 4 through 12.

¹⁰Curtis and Chmelka (1994).

¹¹Schoenbach, Greenleaf, Cziko, and Hurwitz (1999).

¹²Bloome (2001); Dillon, O’Brien, and Volkmann (2001); O’Brien, Moje, and Stewart (2000).

secondary schools means teaching remedial reading, and that teaching reading is the purview of English teachers or reading specialists outside content classrooms.¹³ According to Shanahan, English teachers also do not assume that they should be the ones to teach struggling readers the skills they need.¹⁴ Shanahan further notes the belief of content-area teachers (including English teachers) that if they attempt to teach reading-across-the-curriculum strategies, they will only be taking valuable instructional time away from their designated subject areas.

In short, gaps in the literacy skills of striving adolescent readers and the lack of internal capacity to fill these gaps raise a critical challenge for high school reform initiatives that aim to improve low-performing high schools.¹⁵ These problems are especially acute as students navigate the transition into high school and face a variety of new challenges that can easily push them off the path toward graduation and preparation for postsecondary education and the labor market. Over the past several years, education researchers and practitioners have developed new strategies to address the challenges that ninth-grade students face as they enter high schools, but few have tackled directly the range of problems that arise from limited literacy skills.

Key Elements of a Response and the Role of Supplemental Literacy Programs

In an attempt to mitigate the difficulties that ninth-graders face as they make the transition to high school, many schools are beginning to adopt a range of targeted and comprehensive reform initiatives. Increasingly, these initiatives begin with changes in the structure and organization of the high school through the creation of “smaller learning communities” (SLCs) or even small, independent schools.¹⁶ These structural reforms are often accompanied by curricular and instructional reforms, some of which may be targeted to students who enter high school with limited literacy and math skills.¹⁷ The ERO project builds directly on this precedent by embedding supplemental literacy interventions in “Freshman Academies” — SLCs composed solely of ninth-grade students. To set the context for the ERO study, the following summarizes the roles that SLCs are increasingly playing in high school reform initiatives.

Typically, SLCs function as “schools within schools” characterized by groups of 100 to 200 students who take at least a core set of classes together from interdisciplinary teacher teams. SLCs seek to foster a personalized atmosphere in which students and teachers come to know and trust each other and hold each other to high standards. In Freshman Academies, ninth-

¹³Roe, Stoodt, and Burns (1998).

¹⁴Shanahan (2004).

¹⁵Quint (2006).

¹⁶Abrams and Oxley (2006).

¹⁷Quint (2006).

graders are grouped into a section of the high school building or in an entirely separate building, where they receive extra support from teachers, counselors, and mentors. Several studies suggest that these academies can be effective structures for supporting students as they make the difficult transition from middle school to high school. These studies indicate that SLCs for ninth-grade students can produce increases in attendance, credit accumulation, and on-time promotion to the tenth grade.¹⁸

Despite the growth of SLCs as a central component of high school improvement strategies, high school reformers have increasingly come to acknowledge that changes in instruction and academic supports may be necessary but are insufficient alone to improve the academic performance of struggling students. While formal literacy instruction is not widely practiced in specific content-area classrooms, supplemental reading programs have been developed to respond to the needs of students who have weak literacy skills. Implementing these interventions within SLCs and Freshman Academies can also provide a particularly strong, supportive structural foundation on which to implement and sustain high-quality instructional interventions, such as supplemental literacy programs. Developmental theory suggests that, from both students' and teachers' perspectives, such instructional changes may be more effective when they are mounted within settings that also attend to students' socioemotional needs.¹⁹

Recently, researchers have begun to identify elements of interventions that are designed to address the literacy needs of struggling adolescent readers. At the same time, very few of these elements have been subjected to rigorous evaluations either alone or in combination with one another. Thus, there has been a growing demand for better evidence about what works, for whom, and under what conditions.²⁰ As described below, the elements of these intervention strategies encompass content-related features and the framework for their implementation.

Content-Related Features²¹

- *Motivation and behavior.* Addresses the question, “Why read?” Includes cooperative learning environments and use of high-interest materials.

¹⁸Quint, Miller, Pastor, and Cytron (1999); Kemple and Herlihy (2004); Kemple, Connell, Legters, and Eccles (2006).

¹⁹Kemple, Connell, Legters, and Eccles (2006).

²⁰Alliance for Excellent Education (2004); Alvermann (2002); Biancarosa and Snow (2004); Guthrie and Alvermann (1999); Kamil (2003); National Reading Panel (2000); RAND Reading Study Group (2002); Snow and Biancarosa (2003).

²¹National Reading Panel (2000); Beck, McKeown, and Kucan (2002); RAND Reading Study Group (2002); Snow and Biancarosa (2003); Biancarosa and Snow (2004).

- *Advanced phonics and decoding.* Accounts for the range of expertise in adolescents’ mastery of alphabetic sounds and word decoding. Uses word study that teaches how to decode while simultaneously teaching meaning.
- *Fluency.* Uses guided oral reading at students’ individual reading levels. Includes practice with expository and narrative text.
- *Vocabulary.* Teaches strategies to identify and learn new words and to build context for new words and concepts. Uses both direct and indirect techniques for teaching vocabulary.
- *Comprehension.* Teaches components of text structure, generically and with specific reference to content-area learning. Uses both modeling and instruction to teach strategies and thought processes. Activates students’ prior knowledge and encourages higher-order thinking.
- *Metacognition.* Teaches students to reflect on how they read, to recognize faulty comprehension, and to apply “fix-up” strategies.
- *Writing.* Teaches a process for writing (planning, writing, feedback, editing) that will be successful across the high school curriculum. Promotes use of higher-order thinking skills.

Implementation Framework²²

- *Instructional approach.* Relies on both direct comprehension instruction and student self-directed learning. Includes whole-group, small-group, and individualized instruction. Instruction should be embedded in content and should link concepts, skills, and strategies across topics and over time.
- *Scheduling and duration.* Provides students a minimum of 225 minutes of literacy instruction per week (organized as 45-minute classes each day or as 80- to 90-minute blocked classes every other day), over and above the regular English or language arts classes. Includes lessons or instructional segments that can extend for a full academic year.
- *Group size.* Can accommodate up to 15 students per period to facilitate multiple modes of instruction and attention to individual needs.

²²Biancarosa and Snow (2004).

- *Materials.* Includes diverse reading materials, highly engaging and appropriate for age and skill level.
- *Use of technology.* Uses technology for practice of skills and strategies presented by the teacher.
- *Teacher training and support.* Includes intensive introductory training followed by on-site coaching and ongoing technical assistance. Provides teachers with resources and guides to conduct instruction and assess student progress.
- *Assessment.* Includes regular assessment of reading skills and ties the results to instruction. Uses assessment both to diagnose problems and to monitor progress.
- *Cost.* Must be affordable to allow for adoption by low-income districts.

An array of programs has been developed with one or two of these elements embedded in them.²³ Yet, very little has been done to develop an overall strategy for directing and coordinating a multidimensional response to the needs of students who face the greatest risk of school failure by virtue of their limited literacy skills. In their high-profile call to action to address the needs of struggling adolescent readers, Biancarosa and Snow call for a series of demonstrations that attend to the challenges and variations associated with different components, implementation strategies, and contexts and that are subject to a rigorous assessment of their impact on participating students.²⁴ The ERO study represents a direct and systematic response to this call to action.

Overview of the ERO Study

The ERO study is both a demonstration of two supplemental literacy interventions across a range of contexts and a rigorous evaluation of the interventions' impact on students' reading comprehension skills and their academic performance as they move through high school. The study is a collaboration between policy and research interests that encompass practical responses to important educational problems and a commitment to learning whether these responses produce their desired effects. The U.S. Department of Education's Office of Elementary and Secondary Education (OESE) is providing direct support for implementation to the participating schools and districts, while its Institute of Education Sciences is overseeing the design and execution of the evaluation effort. Incorporating the evaluation expertise of the re-

²³For a summary of the evidence base on interventions that incorporate the elements listed above, see Biancarosa and Snow (2004).

²⁴Biancarosa and Snow (2004), p. 23.

search team, the substantive knowledge of the model developers and the operational capacity of participating sites, the ERO project places a useful policy instrument at the service of both helping students and building knowledge. Following is a brief overview of the demonstration and evaluation components of the ERO study.

A Demonstration of Supplemental Literacy Interventions

The ERO study tracks the implementation of two established supplemental literacy interventions that were developed for high school students whose reading skills are two or more years below grade level as they enter high school. Both programs incorporate many of the design elements discussed above including careful attention to student motivation, a focus on reading fluency, vocabulary, and comprehension, development of metacognition to promote reflective reading strategies, and use of technology. Each program is a full-year course that substitutes for a ninth-grade elective class and is scheduled for a minimum of 225 minutes of instruction per week. They are both designed to accommodate class sizes of 12 to 15 students.

As part of their proposal to participate in the ERO study, the developers of both programs provided suggestive evidence of their developmental appropriateness for the target population of students and of their alignment with the available research base on strategies for improving the literacy skills of struggling adolescent readers.²⁵ Each intervention was part of a larger and more comprehensive high school reform initiative. For the purposes of the ERO study, the programs were modified somewhat and adapted for implementation as an independent class that would replace a regular elective class for ninth-grade students. In order to meet the special needs of high school teachers who do not have reading instruction credentials, the programs' developers also intensified their professional development and coaching strategies. While the two programs share core goals and many instructional strategies, they differ primarily in their approach to implementation.

The supplemental literacy programs are being 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).²⁶ Experienced, full-time English/language arts or social studies teachers volun-

²⁵For an overview of research related to Reading Apprenticeship Academic Literacy, see Schoenbach, Greenleaf, Cziko, and Hurwitz (1999). For an overview of research related to Xtreme Reading and the Strategic Instruction Model, see Schumaker and Deschler (2003, 2004).

²⁶For a complete application package for the special competition, see U.S. Department of Education (2005). The special grant competition was part of OVAE's Smaller Learning Communities initiative and was designed to provide extra funding to qualifying districts for the implementation of the supplemental literacy programs and participation in the ERO evaluation. The grants also included funds for general support of the Smaller Learning Communities initiatives under way in the districts. In 2006, responsibility for the Smaller Learning Communities initiative and for the special ERO grants was moved from OVAE to OESE.

teered to teach the programs for a period of two years. It should be noted that the participating sites were not selected to be representative of all districts and schools across the country. As a result, findings from the ERO study cannot be generalized statistically to the full population of districts and high schools or to urban districts and schools. At the same time, the participating sites reflect much of the diversity of midsize and large urban school districts that serve low-income and disadvantaged populations of students. Thus, the findings will be widely applicable and highly relevant to districts and high schools that are struggling to meet the needs of ninth-graders who lack the literacy skills required for academic success.

A Rigorous Impact Evaluation

The ERO evaluation will unfold over a five-year period and will address the following questions:

- What are the short-term impacts of these supplemental literacy interventions on ninth-grade students' reading skills and behaviors?
- For which subgroups of students are supplemental literacy interventions most or least effective?
- What factors promote or impede successful implementation of the supplemental literacy interventions? In what ways are implementation fidelity and quality associated with program impacts (or lack of impacts) on reading achievement and other outcomes?
- What are the longer-term impacts on other academic outcomes, such as achievement on high-stakes standards-based assessments, performance in academic courses, and progress toward graduation? What is the nature of the relationship between the impacts on reading skills and the impacts on these other outcomes?

The current report provides an early assessment of the first three of these questions as reflected in the first year of implementation. Subsequent reports will provide evidence about the effectiveness of maturing versions of the programs and will address the questions about longer-term impacts.

The ERO evaluation utilizes a two-level random assignment research design. First, within each district, eligible high schools were randomly assigned to use one of the two supplemental literacy programs. This feature of the design allows a direct comparison of the effectiveness of the two programs and avoids confounding the effect of purposeful or self-selection of schools to use the two programs with a true difference in the programs' impact on student achievement.

The second feature of the study design involves the random assignment of eligible and appropriate students within each of the participating high schools. Each high school was asked to identify at least 100 ninth-grade students who were reading at least two years below grade level. Approximately 55 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 enrolled in or continued in a regularly scheduled elective class. This feature of the design is possible because there were more eligible and appropriate students in each high school than the 50 to 60 students that the literacy programs are able to serve. Students in both groups take the regular English/language arts classes offered by their schools as well as other core academic and elective classes required of or offered to ninth-graders. The study includes two cohorts of ninth-grade students: one cohort that was enrolled in the study at the beginning of the 2005-2006 school year and one cohort that was enrolled in the study starting in the 2006-2007 school year.

Finally, the ERO evaluation taps a variety of data sources to measure students' reading achievement and school performance and to assess the fidelity of program implementation.

Overview of This Report

The remaining chapters in this report provide further background on the study design and discuss the implementation and impact findings. Chapter 2 describes the sample of schools and the first cohorts of students who are participating in the study. Chapter 3 presents an in-depth description of the two supplemental literacy programs and their implementation during the initial year of the study. Chapter 4 examines student enrollment and attendance in the ERO classes and looks at the rate at which students in the study's non-ERO sample participated in supplemental literacy services both in and outside school. Chapter 5 reports on the early impacts of the literacy interventions.

This report provides an early look at the implementation and impact of the two literacy interventions based on their initial year of operation in the participating schools. Because of the late award of the special SLC grants, none of the high schools was able to begin its program at the start of the school year. Also, the schools and teachers had no prior experience with the programs, and their knowledge and expertise evolved throughout the year. The delay in program start-up and the schools' and teachers' evolving competence with them means that the interventions did not receive as complete a test as would be expected with a full year of operation and prior experience with implementation. As a result, the findings presented in this report should be interpreted cautiously in terms of their implications for education policy and practice. Later reports on the ERO evaluation will provide more conclusive evidence about the effectiveness of these interventions and a more solid footing for use by policymakers and practitioners. Despite the limitations of an early assessment of program experiences, the current report aims to offer useful insights into the characteristics, implementation, and impact trends of these interventions.

Chapter 2

Study Sample and Design

This chapter describes the sample of schools and students involved in the Enhanced Reading Opportunities (ERO) study, the different sources of data and the impact measures created from these data, student response rates during follow-up data collection, and the analytic methods used to assess program impacts. The chapter discusses the following key points:

- Thirty-four schools from 10 school districts were selected for the study and were randomly assigned to use one of the two supplemental literacy programs. The resulting two groups were similar on a range of school characteristics.
- The study sample includes 2,916 students with baseline reading test scores that fell between two and five years below grade level. Fifty-seven percent of these students were randomly assigned to the ERO group and were scheduled into the ERO classes, and the remaining 43 percent were assigned to a non-ERO control group and continued in a ninth-grade elective class.
- Approximately 83 percent of the students in the study sample (a total of 2,413 students) completed the follow-up reading assessment and survey. Among respondents, overall differences found in background characteristics between the ERO and non-ERO groups are not statistically significant.
- Statistical-power calculations indicate that the full study sample available for the impact analysis is sufficient for minimum detectable effects sizes of 0.06 standard deviation units or larger for the reading test score outcomes. The samples available for each of the two supplemental literacy programs are sufficient for minimum detectable effects sizes of 0.10 standard deviation units or larger.

School Sample

The school districts participating in this study were selected through a special grant competition run by the Office of Vocational and Adult Education (OVAE) within the U.S. Department of Education (ED).¹ As an extension of the Smaller Learning Communities (SLCs) grant program, this competition sought to provide funding for the implementation of two sup-

¹U.S. Department of Education (2005).

plemental ninth-grade literacy programs in selected high schools and to sustain and enhance existing SLCs in these high schools.

In June 2005, ED selected 10 grantee school districts encompassing 34 high schools — from a pool of 33 applicant districts.² The 10 grantee districts encompass 65 high schools, with the smallest district having four high schools and the largest having 22 high schools. Seven of the grantee districts included four of their high schools in the study, and the remaining three districts included two high schools. Grantee districts will receive approximately \$1.25 million over five years for each participating high school. From their SLC grants, districts were required to set aside \$250,000 per high school over the first two years of their grant period to cover the costs of implementing the supplemental reading programs, including costs associated with teachers' salaries and benefits, teacher-training activities, coaching and materials to be provided by the program developers, classroom computers, and other equipment and materials.

Random Assignment of Schools

Following the selection of grantee districts to participate in the ERO study, the study team randomly assigned the participating schools to implement one of the two literacy programs. Within each district, half the participating schools were randomly assigned to the Reading Apprenticeship Academic Literacy program, and half were randomly assigned to Xtreme Reading. Schools were randomly assigned to the interventions as a safeguard against selection bias. That is, if districts and developers had been allowed to choose the allocation of the interventions, the potential would have existed for decisions to have been made based on any of a variety of characteristics associated with outcomes of overall effectiveness that might have made one school a more favorable candidate over another for a more “successful” implementation of the program. Such characteristics cannot be measured, thereby presenting a possible threat to the validity of the study. Essentially, by randomly assigning schools to one of the two supplemental literacy interventions, the study ensured that the intervention developers could not select schools that were higher performing or at a higher level of readiness for their programs. It also ensured that the schools could not select a literacy program that they believed would be more appropriate or more effective for their school. As a result, differences in impacts that may emerge between the two groups of schools can be attributed to differences between the two programs rather than to differences in school characteristics or the method for assigning schools to the programs.

²The number of applicants for the special SLC Grant Competition was reported to the study team by OVAE staff.

Characteristics of Schools Selected for the ERO Project

Table 2.1 presents characteristics of the 34 high schools participating in the ERO study. Overall, ERO programs were implemented in schools located predominantly in large and mid-size cities, with some of the schools in each of these categories being listed as “urban fringe.” As specified by the OVAE grant requirements, all schools enrolled more than 1,000 students in grades 9 through 12, averaging 1,685 students per school. The schools enrolled an average of 570 ninth-grade students, ranging from 320 to 939 ninth-grade students per school. Table 2.1 shows the average “promoting power” for the participating schools, which can serve as a proxy for the likely longitudinal graduation rate.³ It indicates that the twelfth-grade class is 59 percent of the size of the ninth-grade class three years earlier, suggesting that roughly 41 percent of students have left the schools between the ninth and twelfth grades. The table also shows that 38 percent of the students in the participating schools were eligible for Title I services and that 47 percent of the students were approved for free or reduced-price lunch.

Overall, Table 2.1 indicates that there is a high degree of similarity between the schools randomly assigned to use Reading Apprenticeship Academic Literacy and the schools assigned to use Xtreme Reading.

Table 2.1 also includes information about all high schools across the country that, like those selected for the ERO study, are located in large and midsize cities, served over 1,000 students in grades 9 through 12, and did not select students based on past achievement or performance. This national census of similarly situated high schools provides a reference point that helps contextualize and describe the ERO high schools. In comparison with the national sample, the schools selected for the ERO study include a higher proportion of students with characteristics associated with low performance. The ERO schools have lower levels of student promotion, higher percentages of students eligible for free and reduced-price lunch, and higher eligibility for Title I funding. Additionally, the populations at ERO schools comprise higher percentages of minority students than the national sample.

Student Sample

At the inception of the ERO project, the primary target population for the supplemental literacy interventions included students entering ninth grade with reading skills that were between two and four years below grade level. To qualify for an ERO grant, districts were re-

³Balfanz and Legters (2004) developed this measure of “promoting power” to approximate a school’s graduation rate. It is calculated as the ratio of the number of twelfth-grade students in a given school year to the number of ninth-grade students from three years prior.

The Enhanced Reading Opportunities Study

Table 2.1

**Characteristics of ERO Schools and Average Schools
in the United States (2004-2005)**

Characteristic	All ERO Schools	Reading Apprenticeship Schools	Xtreme Reading Schools	Average U.S. Schools ^a
Average number of students	1,685	1,687	1,683	1,866
Average number of students in grade 9	570	566	574	556
Average number of students in grade 10	432	436	429	478
Average number of students in grade 11	358	359	358	424
Average number of students in grade 12	317	312	322	382
Average promoting power ^b (%)	59.1	56.7	61.6	75.4
Students eligible for free or reduced-price lunch (%)	46.9	44.5	49.2	30.0
Race/ethnicity (%)				
Hispanic	25.1	24.6	25.6	19.3
Black	41.1	41.9	40.4	19.7
White	31.2	31.0	31.5	53.5
Other	2.6	2.6	2.6	7.0
Eligible for Title I (%)	38.2	41.2	35.3	26.0
Locale (%)				
Large city ^c	52.9	52.9	52.9	61.2
Midsize city ^d	47.1	47.1	47.1	38.8
Sample size	34	17	17	3,727

SOURCES: U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD), "Public Elementary/Secondary School Universe Survey Data", 2004-2005 and 2001-2002.

NOTES: This table provides information on 34 ERO schools from 10 districts.

Rounding may cause slight discrepancies in calculating sums and differences.

^a"Average U.S. Schools" includes schools that have more than 1,000 total students, have more than 100 students in each grade during 2004-2005, have at least 125 students in the ninth grade during 2001-2002, are noncharter schools, are located in a large or midsize city or in the urban fringe of a large or midsize city, are defined as "regular" schools by the Common Core of Data, and are operational at the time of the Common Core of Data report.

^b"Promoting power" is calculated as the ratio of twelfth-grade students in 2004-2005 to ninth-grade students in 2001-2002.

^c"Large city" is defined as a city having a population greater than or equal to 250,000. Schools in this category also include the urban fringe of a large city.

^d"Midsize" city is defined as a city having a population less than 250,000 but greater than 50,000. Schools in this category also include the urban fringe of a midsize city.

quired to provide documentation that each high school would include at least 125 ninth-grade students with reading skills at these levels.⁴

Among the first tasks for the ERO study were to identify potentially eligible students in each of the participating high schools, obtain parental consent for the students to be included in the study sample, and administer a baseline reading test and a survey. Then, assuming that 125 students were eligible for the ERO programs and consented to be in the study, the study team would conduct random assignment such that up to 60 of these eligible students would be selected to enroll in the ERO classes. Of the students randomly assigned to the ERO program, the school was responsible for scheduling those students into four ERO class sections. Typically, those sections each contained 12 to 15 students. Of the remaining 65 students, up to 50 students would be assigned to enroll or remain in a regular ninth-grade elective class. The remaining 15 students would constitute a nonresearch waiting list and would be admitted to an ERO class if enrollment levels fell below the desired minimum of 12 students, due to attrition over the school year.⁵ Because the special SLC grants were not awarded until the summer of 2005, this process could not begin until the start of the 2005-2006 school year. This meant that the student study sample would not be identified until several weeks into the school year and that students selected for the ERO classes would be forced to withdraw from an elective course they had already begun to attend.

Early in the 2005-2006 school year, it became clear that the study team and the schools were facing significant challenges that would require some modification in the original targeting criteria and that would delay the start of the classes further. The study team was in regular contact, both in person and by telephone, with staff in the participating schools and districts to monitor the student testing and recruitment process. The team learned that several of the schools had fewer than the prescribed number of students in the target range — at least according to the reading test that was being used for the ERO study. Also, all the schools faced severe challenges in getting eligible students to return signed consent forms. As a result, the study sample was expanded to include students between two and five years below grade level, and the eligibility criteria to be in an ERO class were expanded to include students with reading levels between one and five years below grade level. Schools also employed more intensive strategies to obtain consent forms. In the end, all the participating schools were able to meet minimal targets for the study sample, but this was not completed until an average of six weeks into the school year.

⁴It should be noted that English Language Learning (ELL) and special education students who required specific classroom, instructional, or testing accommodations were not eligible for the ERO classes. The ERO programs were not designed to accommodate the special needs of these students nor the potential scheduling conflicts with other services that the students were likely to receive.

⁵Note that students assigned to the nonresearch waiting list were not included in the analysis, even if they were later scheduled into ERO classes.

Following a more detailed discussion of the student recruitment and random assignment process, this section of the chapter describes the characteristics of the core sample of students in the study's first cohort.

Student-Level Random Assignment

Because the special SLC grants were not awarded until the summer of 2005, the student recruitment process did not begin until the start of the 2005-2006 school year. Staff from each of the 34 high schools administered the Group Reading Assessment and Diagnostic Examination (GRADE) to their ninth-grade students. Students who scored between the fourth- and eighth-grade level on the GRADE reading comprehension subtests were considered eligible for the ERO classes. Eligible students were then asked to return a parental consent form giving permission to participate in the study and to enroll in the ERO classes if they were selected. Once eligible students returned a signed affirmative consent form and completed the baseline survey, they were entered into MDRC's random assignment database. While the recruitment of eligible students required the assistance of school and district staff members in communicating with parents and students and collecting consent forms, computerized random assignment of students was conducted solely by MDRC staff.

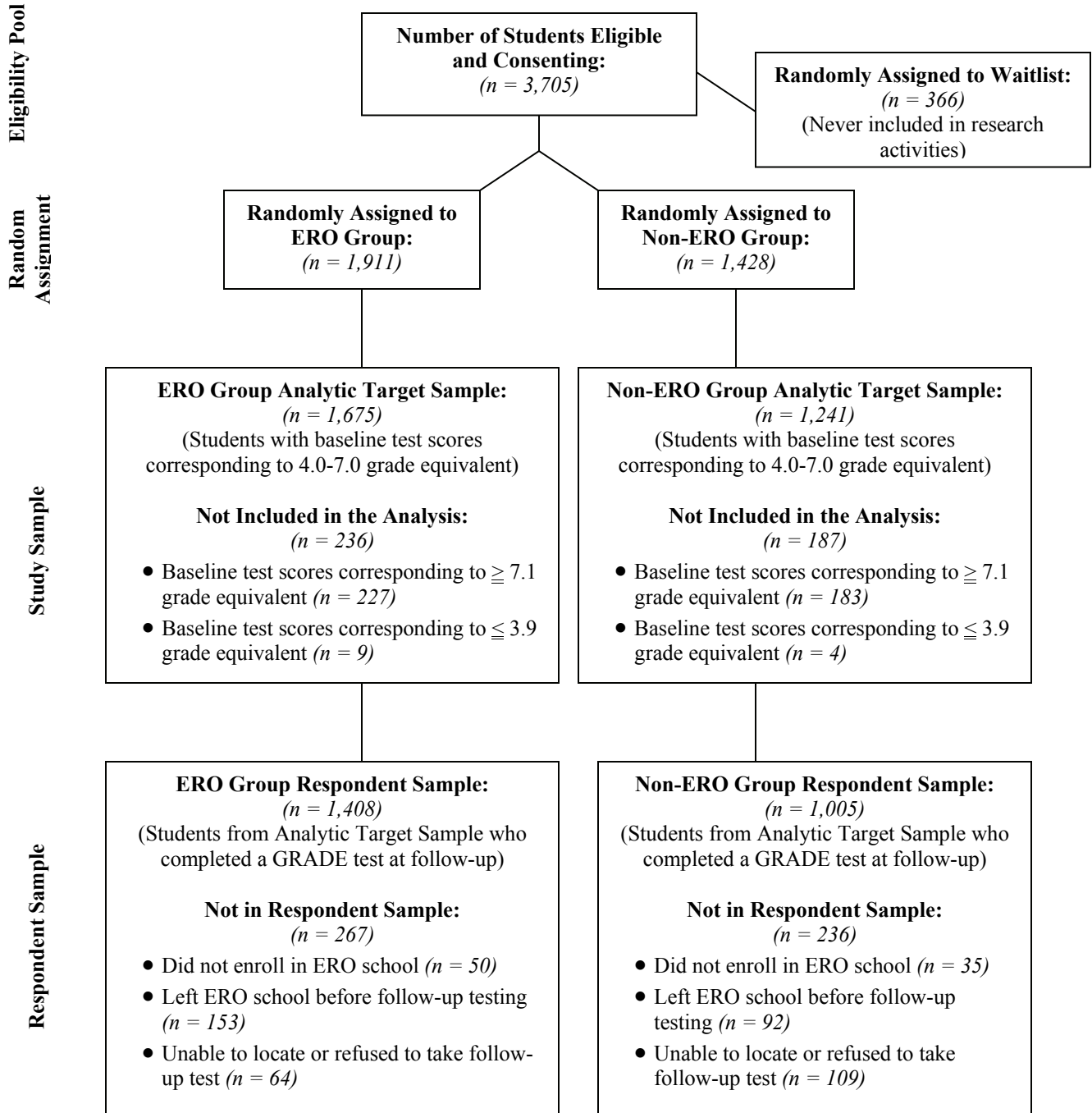
The ERO programs were designed to accommodate between 12 and 15 students per class, and each high school was required to offer four ERO class sections. The study team identified 3,339 eligible and consenting students from across the 34 participating high schools (on average, 98 students per school). Figure 2.1 shows that 1,911 (57 percent) of these students were randomly selected to enroll in the ERO classes (referred to as the "ERO group") and 1,428 (43 percent) were randomly assigned to the control group (referred to as the "non-ERO group"). Although the eligibility criteria were expanded to include students with test scores ranging from the 4.0 to 8.0 grade equivalent to keep the classes at capacity, the analyses in this report focus exclusively on the students whose baseline test scores ranged from the 4.0 to 7.0 grade equivalent (two to five years below the ninth-grade level). Figure 2.1 shows that there are 2,916 students in this group (87 percent of the entire study sample; on average, 86 students per school), with 1,675 (57 percent) randomly assigned to the ERO group and 1,241 (43 percent) randomly assigned to the non-ERO group.⁶ All further references in this report to the "study sample" refer to students with scores ranging from the 4.0 to 7.0 grade equivalent.

⁶A total of 410 students had scores that were equivalent to the 7.1 grade equivalent or higher. In addition, 13 students had scores that were equivalent to the 3.9 grade equivalent or lower. Given that the two interventions and the evaluation were designed primarily to test the effects of supplementary literacy interventions on ninth-grade students with reading comprehension skills between the fourth- and seventh-grade levels, data for these 423 students are not included in the impact analysis for this report.

The Enhanced Reading Opportunities Study

Figure 2.1

Construction of the Impact Sample from the Eligibility Pool



Characteristics of the Study Sample

The background characteristics of the ERO group and the non-ERO group were compared to determine whether random assignment resulted in two equivalent groups. There is a high degree of similarity between the two groups' baseline characteristics, as illustrated in Table 2.2. On average, students in the study sample had a reading comprehension composite score of just under 86 standard score points on the GRADE reading assessment. This average corresponds to the 5.1 grade level (an average of almost four years below grade level at the beginning of ninth grade) and to the 16th percentile nationally. The study sample is over 70 percent Hispanic or black; about 45 percent of the students speak a language other than English at home; and about 30 percent are overage for grade (15 years old or older at the start of ninth grade, suggesting that they were retained in a prior year).⁷ A general F-test indicates that, overall, there are no systematic differences in these characteristics between the ERO and non-ERO groups in the study sample. The lack of systematic differences indicates that random assignment was successful in creating two equivalent research groups at baseline. Similar results were found when examining the background characteristics of study-sample students from the Reading Apprenticeship sites and the Xtreme Reading sites, separately.⁸

Data Sources and Measures

The ERO evaluation utilizes a variety of data sources to measure students' reading achievement and reading behaviors and to assess the fidelity and quality of program implementation. Following is an overview of the data sources utilized in the current report.

Group Reading Assessment and Diagnostic Examination (GRADE)

The GRADE is a norm-referenced, research-based reading assessment that can be administered to groups. It is meant to be a diagnostic tool to assess what reading skills individuals have and what skills need to be taught.⁹ It is used widely to measure performance and track growth of an individual student and groups of students from fall to spring and from year to year. The GRADE contains multiple subtests, including two reading comprehension subtests (sentence comprehension and passage comprehension), a listening comprehension subtest, and a vocabulary subtest. For the ERO study, the two reading comprehension subtests (Level H, Form A) were administered to all students prior to random assignment. Near the end of their ninth-grade year, students completed the two reading comprehension subtests (Level H, Form B) as

⁷National Center for Education Statistics (1990).

⁸See Appendix B.

⁹See American Guidance Service (2001a, 2001b) for technical information about the GRADE.

The Enhanced Reading Opportunities Study

Table 2.2

**Characteristics of Students in Cohort 1
Full Study Sample**

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Race/ethnicity (%)				
Hispanic	31.8	31.7	0.1	0.943
Black, non-Hispanic	44.6	45.4	-0.8	0.556
White, non-Hispanic	17.7	17.0	0.7	0.585
Other	5.9	5.8	0.1	0.940
Gender (%)				
Male	49.9	50.1	-0.3	0.878
Female	50.1	49.9	0.3	0.878
Average age (years)	14.8	14.8	0.0	0.152
Overage for grade ^a (%)	31.5	28.3	3.1	0.054
Language other than English spoken at home (%)	45.6	45.5	0.1	0.974
Language spoken at home missing (%)	6.7	6.8	-0.1	0.921
Mother's education level (%)				
Did not finish high school	18.1	19.0	-0.8	0.554
High school diploma or GED certificate	25.0	24.8	0.1	0.942
Completed some postsecondary education	29.3	30.2	-0.9	0.581
Don't know	20.2	18.8	1.3	0.360
Missing	7.4	7.1	0.3	0.728
Father's education level (%)				
Did not finish high school	16.8	17.9	-1.1	0.444
High school diploma or GED certificate	23.0	23.2	-0.2	0.899
Completed some postsecondary education	18.3	20.6	-2.4	0.104
Don't know	33.6	29.8	3.8 *	0.027
Missing	8.3	8.5	-0.2	0.825
GRADE reading comprehension^b				
Average standard score	85.7	86.1	-0.3	0.093
<i>Corresponding grade equivalent</i>	5.1	5.2		
<i>Corresponding percentile</i>	16	17		
6.0 - 7.0 grade equivalent (%)	33.2	35.8	-2.6	0.140
5.0 - 5.9 grade equivalent (%)	29.6	27.6	1.9	0.251
4.0 - 4.9 grade equivalent (%)	37.3	36.6	0.7	0.695
Sample size	1,675	1,241		

(continued)

Table 2.2 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year.

The differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is the ERO group value minus the difference.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

^bThe 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 A). No statistical tests or arithmetic operations were performed on these reference points.

well as the vocabulary subtest. In addition to the raw score (the total number of items answered correctly), the GRADE also provides standardized scale, normal curve equivalent, grade equivalent, percentile, and stanine scores.

The primary measure of reading achievement for this study is students' scores on the GRADE reading comprehension assessment. This component of the GRADE includes subtests that measure sentence comprehension and passage comprehension. According to the GRADE technical manual, "the purpose of sentence comprehension is to identify if the student can comprehend a sentence as a whole thought or unit."¹⁰ The GRADE technical manual also characterizes passage comprehension as measuring a student's skills in understanding an extended passage consisting of a single paragraph or multiple paragraphs.¹¹ A central objective of each of the two ERO programs is to provide students with immediate and intensive instruction in the use of strategies and skills that expert readers use to understand written texts. Thus, for the purposes of the ERO evaluation, the GRADE reading comprehension assessment serves as the primary early indicator of the programs' effectiveness.

A secondary measure of students' reading achievement is their scores on the GRADE vocabulary assessment. According to the GRADE technical manual, the vocabulary subtest is intended to measure a student's knowledge of word meanings with minimal contextual clues.¹² Each of the two ERO programs provides some instruction aimed at helping students break down word meanings through advanced decoding skills and strategies for recognizing word structures (root words, prefixes, and suffixes). Thus, the GRADE vocabulary assessment can

¹⁰American Guidance Service (2001a), p. 39.

¹¹American Guidance Service (2001a), p. 45.

¹²American Guidance Service (2001a), p. 45.

provide indication of whether these approaches increase the stock of words that students know. However, because the two ERO programs focus primarily on helping students use contextual clues to understand the meaning of words, the vocabulary subtest is seen as a secondary indicator of the programs' effectiveness.

The GRADE reading comprehension and vocabulary performance levels and impacts for the ERO and non-ERO groups are presented in standard score units provided by the American Guidance Service, which publishes the GRADE.¹³ Standard scores are a more accurate representation of a student's level of performance than raw scores because they have uniform meaning from one test period to another and from one grade level to another. Standard scores indicate how far a student's performance on the test is from the average for all students at a given grade level, and standard scores take into account the variability of scores among a nationally representative group of students in that grade. Also, standard scores on the GRADE can be compared with standard scores on other tests of reading comprehension and vocabulary.

To help the reader interpret the standard score values, impact tables also present the national grade equivalent and national percentile that correspond most closely to the average standard score for the ERO and non-ERO groups, respectively. A grade equivalent score is the grade at which a particular raw score or standard score represents the median for the test's norming population. For example, a grade equivalent score of 9.0 refers to a median performance at the beginning of ninth grade, and a 9.8 grade equivalent indicates a median performance at the end of ninth grade.¹⁴

The reading comprehension and vocabulary test score impact estimates are presented both in standard score units and in effect-size units. Effect sizes provide an indication of the magnitude of the impact estimates relative to the overall variation in test scores for students in

¹³Specifically, each student's raw scores on the GRADE subtests and composite scores were converted to standard scores based on national norms for Level H, Grade 9, Spring Testing (American Guidance Service, 2001b, pp. 30-33). Based on these norms, 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. A standard score of 85 corresponds, approximately, to the 4.9 grade equivalent.

¹⁴Note that grade equivalents and percentiles are not equal-interval scales of measurement. Grade equivalents indicate a student's place along a growth continuum, which may not increase at regular intervals. For example, the difference between a vocabulary grade equivalent of 1.0 and 2.0 represents a greater difference in vocabulary knowledge than the difference between a grade equivalent of 8.0 and 9.0. Percentiles indicate the percentage of students in the test's norming group who performed at or below a given student's score. As such, percentiles provide information only about the rank order of students' scores; they do not provide any information about students' actual performance. Because they do not reflect equal intervals between units of measure, neither grade equivalents nor percentiles can be manipulated arithmetically. (See American Guidance Service, 2001a, pp. 55-60.) Thus, readers should exercise caution when interpreting differences in grade equivalents or percentiles between the ERO and non-ERO groups and between the baseline and follow-up tests.

the study sample. For the purposes of the impact analysis, effect sizes are calculated as a proportion of the standard deviation of the test scores for students in the non-ERO group at the end of ninth grade.¹⁵ The standard deviation for the non-ERO group reflects the expected variability in test scores that one would find in the absence of the ERO programs. The impact effect size, therefore, provides an indication of how much the ERO programs moved students along this variability in expected performance.

Student Surveys

Students in the study sample completed the Enhanced Reading Opportunities Student Baseline Survey prior to random assignment. The baseline survey includes the following information for students in the study sample: gender, race/ethnicity, age, and current high school. These data items were required for random assignment and are available for all students in the study sample. The baseline survey also includes additional background information and information about students' reading behaviors and attitudes.

The study team administered the Enhanced Reading Opportunities Student Follow-Up Survey to students in the study sample at the same time as the follow-up GRADE assessment. The impact analysis presented in Chapter 5 focuses on three measures of students' reading behavior that were derived from the survey.¹⁶

Each of the ERO programs aims explicitly to increase the amount of time that students spend reading, both for school and for their own enrichment outside school. The programs do this directly by assigning students reading activities during class and for homework. They also attempt to build students' reading skills, confidence, and enjoyment, in the hope that they will take the initiative to read more frequently and for longer periods of time on their own. The first two measures in the reading behaviors impact analysis focus on how often students read various types of texts for school and outside school. Though self-reported by students, these outcomes provide a direct indication of whether the ERO programs are increasing the amount of time that students spend reading.

Amount of School-Related Reading

This measure was constructed to reflect the self-reported number of times during the prior month that a student read each of seven different types of text for school — in school or for homework: history, science, or math textbooks; literary texts; research or technical reports;

¹⁵The standard deviation of the reading comprehension standard score for the non-ERO group at follow-up is 10.458. The standard deviation of the vocabulary standard score for the non-ERO group is 10.505.

¹⁶A list of the survey items used to create these three measures and a copy of the survey instrument are presented in Appendix A.

newspaper or magazine articles; or workbooks. For the purposes of this analysis, the measure assumes that there was an average of 30 days in the prior month and that a student's report of having read each type of text represents a separate reading occurrence. Thus, the measure was constructed to allow for up to 210 self-reported school-related occurrences of reading activities during the prior month (7 survey items; Cronbach's alpha = .83).¹⁷

Amount of Non-School-Related Reading

This measure was constructed to reflect the self-reported number of times during the prior month that a student read each of seven different types of text outside school: fictional books; plays; poetry; (auto)biographies; books about science, technology, or history; newspaper or magazine articles; or reference books. For the purposes of this analysis, the measure assumes that there was an average of 30 days in the prior month and that a student's report of having read each type of text represents a separate reading occurrence. Thus, the measure was constructed to allow for up to 210 self-reported occurrences of reading activities outside school during the prior month (7 survey items; Cronbach's alpha = .73).

The third measure is intended to provide an indication of whether students use some of the skills and techniques that the ERO programs try to teach (asking questions of the text and reviewing and rereading). These strategies are second nature for proficient readers, and the measure can serve as a useful indicator of whether students are starting to incorporate them more explicitly into their reading behavior.

Use of Reflective Reading Strategies

This measure captures students' reported use of reflective reading strategies (each item is rated on a scale from 1 to 4) as they read for their English/language arts class and for one other academic class.¹⁸ Students were asked to rate their use of these two strategies on a scale from 1 (strongly disagree) to 4 (strongly agree) (4 survey items; Cronbach's alpha = .88).

The impact estimates for each of these three measures of reading behavior (amount of school-related reading, amount of non-school-related reading, and use of reflective reading strategies) are presented both in their original metrics and in effect-size units. Effect sizes provide an indication of the magnitude of the impact estimates relative to the variation in the measures for students in the study sample who were not exposed to the ERO programs. As with the

¹⁷Cronbach's coefficient alpha is a statistical measure of the degree to which the individual items used to create the multi-item construct are correlated with each other (Cronbach, 1951).

¹⁸The follow-up survey asked students to report on reading strategies that they use in social studies, science, and mathematics classes, if they are taking these courses. The measure relied on the social studies class, if the student reported taking social studies. Otherwise, it includes science. If the student was not taking either social studies or science, the measure includes mathematics.

test score outcomes, effect sizes are calculated as a proportion of the standard deviation of the given outcome for students in the non-ERO group.¹⁹ The standard deviation for the non-ERO group reflects the expected variability in the reading behavior that one would find in the absence of the ERO programs. The impact effect size, therefore, provides an indication of how much the ERO programs moved students along this variability in expected reading behavior.

Teacher Survey

The study team administered a two-part survey to ERO teachers during the summer training institutes held by the interventions' developers. Part 1 of the survey asked teachers about their backgrounds, their experiences with professional development activities, their school environments, and their beliefs about literacy instruction. Part 2 of the survey asked teachers about their impressions of the training they attended.

Implementation Data

Classroom Observations

The analysis of ERO program implementation fidelity in the first year of the study is based on field research visits to each of the 34 high schools during the second semester of the 2005-2006 school year. The primary data collection instrument for the site visits was a set of protocols for classroom observations and interviews with the ERO teachers.²⁰ The observation protocols provided a structured process for trained classroom observers to rate characteristics of the ERO classroom learning environments and the ERO teachers' instructional strategies. Each of these characteristics were selected for assessment because they were aligned with program elements specified by the developers and, by design, were aligned with supplemental literacy program elements that are believed to characterize high-quality interventions for struggling adolescent readers.²¹ Chapter 3 provides a more detailed description of the data collection process and a description of the summary measures of implementation fidelity that were developed from the classroom observation data. Appendix D provides further background on the properties of the classroom observation data and the fidelity measures.

¹⁹The standard deviation of the "amount of school-related reading" for the non-ERO group is 43.867. The standard deviation of the "amount of non-school-related reading" for the non-ERO group is 31.834. The standard deviation of the "use of reflective reading strategies" for the non-ERO group is 0.670.

²⁰The observation protocols can be found in Appendix D.

²¹Biancarosa and Snow (2004).

Teacher Interviews

During the field visits, the study team interviewed the ERO teacher using a semi-structured interview protocol that focused on teachers' perceptions of aspects of the intervention, of the coaching and support that they received from the developers, of the ease of implementing the program, and of students' responses to and challenges with the program. The study team also interviewed English/language arts teachers and elective teachers in order to explore the extent to which literacy instruction may be taking place in classes other than ERO.

Interviews with District Coordinators

The study team interviewed the ERO district coordinators during the site visits, to gather information as to their perceptions about implementing the program.

ERO Class Attendance Records

Each of the ERO teachers provided monthly school attendance data for all students in the study sample and ERO class attendance data for those students assigned to an ERO class.

Student Course Schedules

Each school provided the study team with copies of the schedules for all students in the study sample. One purpose of the schedule data is to confirm that ERO students were enrolled in the ERO classes and that non-ERO students were not.²² These data allow the study team to check for possible contamination — that is, for non-ERO students receiving the ERO program.

Follow-Up Data Collection and Response Rates

The follow-up GRADE assessment and survey were administered to students in the study sample late in the 2005-2006 school year. Overall, the follow-up data are available for 83 percent of the study sample. Table 2.3 shows that the response rate for students in the ERO group is 84 percent, compared with 81 percent for the non-ERO group. This difference is statistically significant (p-value less than or equal to 5 percent). Although the response rates for students in the ERO groups are similar for both the Reading Apprenticeship and the Xtreme Reading schools, the rate is somewhat lower for students in the non-ERO group from the Reading Apprenticeship schools. The difference in response rates between the ERO and non-ERO

²²See Chapter 4 for discussion of student schedules and enrollment in the ERO classes.

The Enhanced Reading Opportunities Study

Table 2.3

**Response Rates of Students in Cohort 1
Full Study Sample**

	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
<u>All schools</u>				
Response rate (%)	84.1	81.1	2.9 *	0.037
Sample size	1,675	1,241		
<u>Reading Apprenticeship schools</u>				
Response rate (%)	84.6	79.3	5.2 *	0.011
Sample size	811	574		
<u>Xtreme Reading schools</u>				
Response rate (%)	83.6	82.7	0.9	0.649
Sample size	864	667		

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

NOTES: This table represents the response rates for the follow-up GRADE assessment, which was administered in spring 2006 at the end of students' ninth-grade year. The follow-up student questionnaire was also administered at that time. The difference in response rates between the test and survey is negligible.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. The p-value is the probability that the observed difference is the result of chance and does not represent a true difference between groups. The lower the p-value, the less confidence that there is not a difference between the two groups. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

groups in the Reading Apprenticeship schools is statistically significant (p-value less than or equal to 5 percent).²³

When response rates are less than 100 percent or when there are differences between program and control groups, it is important to investigate two concerns. First, does the respondent sample differ from the full study sample and from the nonrespondent sample? Second, within the respondent sample, are the ERO group and the non-ERO group still equivalent?

The ERO study team conducted a nonresponse analysis by examining differences in background characteristics between respondents and nonrespondents in the study

²³See Appendix Table B.1 in Appendix B.

sample.²⁴ While the respondent sample reflects the general characteristics of the full study sample, an overall F-test comparing the respondents and nonrespondents indicates that there are systematic differences between them in student characteristics. Most notably, response rates are lower for students with characteristics associated with doing poorly in school. For example, response rates are lower among students who were overage for grade than for those students who were not likely to have been held back in a previous grade. There are also differences in response rates across the participating high schools. Overall, however, response rates are similar for the schools using the Reading Apprenticeship program (82 percent) and those using Xtreme Reading (83 percent). The overall differences between respondents and nonrespondents suggest that one should be cautious when generalizing findings from the first cohort follow-up respondent sample.²⁵

As noted earlier, the three percentage point difference in the response rates between the ERO group (84 percent) and the non-ERO group (81 percent) is statistically significant (p-value less than or equal to 5 percent). This raises a concern about whether respondents in the ERO group differ systematically from respondents in the non-ERO group. Table 2.4 shows the background characteristics of all 2,413 students in the first cohort follow-up respondent sample and provides a comparison between the ERO and non-ERO groups. Like Table 2.2 for the overall study sample, Table 2.4 shows a high degree of similarity between the respondents in the ERO and non-ERO groups across the baseline characteristics. A general F-test indicates that, overall, there are no systematic differences between the ERO and non-ERO group respondents.²⁶ This suggests that one may have a high degree of confidence that differences in outcomes between the two groups reflect impacts of the ERO programs rather than preexisting differences in background characteristics.

The characteristics displayed in Table 2.4 indicate that the typical follow-up respondent sample member was reading well below grade level at the start of ninth grade and that many students have characteristics associated with a risk of doing poorly in school. On average, students had the same reading comprehension composite score of about 86 standard score points, corresponding to the 5.2 grade level and to the 17th percentile nationally. Also, over 70 percent of the students in the follow-up respondent sample are Hispanic or black, and over 45 percent

²⁴See Appendix B for the results of the statistical analyses that were conducted to assess differences between respondents and nonrespondents. Results are presented for all the participating high schools together and, separately, for the groups of schools using Reading Apprenticeship and Xtreme Reading, respectively.

²⁵See Appendix F for results from supplemental impact analyses that include sampling weights to account for differences between respondents and nonrespondents. These results indicate very little difference between the weighted and unweighted impact estimates.

²⁶See Appendix B for the results of the statistical analyses that were conducted to assess differences between the ERO and non-ERO groups in the respondent sample.

The Enhanced Reading Opportunities Study

Table 2.4

Characteristics of Students in Cohort 1
Follow-Up Respondent Sample

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Race/ethnicity (%)				
Hispanic	32.7	33.0	-0.3	0.803
Black, non-Hispanic	42.9	43.6	-0.7	0.632
White, non-Hispanic	18.3	17.2	1.1	0.437
Other	6.2	6.2	0.0	0.999
Gender (%)				
Male	50.1	51.3	-1.3	0.542
Female	49.9	48.7	1.3	0.542
Average age (years)	14.8	14.7	0.0	0.103
Overage for grade ^a (%)	28.1	25.1	2.9	0.092
Language other than English spoken at home (%)	47.1	45.9	1.2	0.512
Language spoken at home missing (%)	6.7	7.2	-0.5	0.618
Mother's education level (%)				
Did not finish high school	17.0	16.8	0.2	0.891
High school diploma or GED certificate	25.5	24.7	0.8	0.641
Completed some postsecondary education	29.0	31.3	-2.2	0.229
Don't know	21.0	19.7	1.3	0.426
Missing	7.5	7.7	-0.1	0.885
Father's education level (%)				
Did not finish high school	16.7	16.9	-0.2	0.894
High school diploma or GED certificate	22.9	22.1	0.8	0.645
Completed some postsecondary education	18.2	22.1	-3.9 *	0.015
Don't know	33.9	30.0	4.0 *	0.038
Missing	8.2	8.9	-0.7	0.518
GRADE reading comprehension ^b				
Average standard score	85.9	86.2	-0.3	0.143
<i>Corresponding grade equivalent</i>	5.1	5.2		
<i>Corresponding percentile</i>	16	17		
6.0 - 7.0 grade equivalent (%)	34.4	37.0	-2.6	0.193
5.0 - 5.9 grade equivalent (%)	29.3	26.4	2.9	0.115
4.0 - 4.9 grade equivalent (%)	36.2	36.6	-0.3	0.859
Sample size	1,408	1,005		

(continued)

Table 2.4 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year.

The differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is the ERO group value minus the difference.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

^bThe 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 A). No statistical tests or arithmetic operations were performed on these reference points.

reported that a language other than English is spoken in their homes. Tables 2.5 and 2.6 present similar results for students in the follow-up respondent samples from the Reading Apprenticeship schools and from the Xtreme Reading schools, respectively.

The similarity between the student characteristics of the follow-up respondent sample and full study sample — as well as the lack of systematic differences between the ERO and non-ERO groups in the follow-up respondent sample — indicate that the follow-up respondent sample preserves the balance that was achieved with random assignment for the full study sample. This balance was also preserved in the groups of schools using each of the two supplemental literacy programs.

Analytic Methods and Procedures

When examining the effectiveness of the ERO programs in improving students' reading achievement and behaviors, it is important to distinguish between measures of program “outcomes” and measures of program “impacts.” *Outcomes* refer to the measures of student performance, behaviors, achievement, and attitudes — in this case, reading achievement and reading behaviors at the end of the ninth-grade year. An *impact* is the effect that the ERO programs have on an outcome. The average outcome levels for students in the ERO group alone provide potentially misleading conclusions. Reading achievement and behaviors are likely to change for students for reasons not related to a special intervention like the ERO programs. In order to determine the net effect, or “value added,” of the ERO programs, it is necessary to compare the experiences of a group of students who were exposed to the ERO classes with a similar group of students who also applied but were not selected to enroll. As discussed earlier in this chapter, the ERO and non-ERO groups participating in this study were determined through a random

The Enhanced Reading Opportunities Study

Table 2.5

Characteristics of Students in Cohort 1
Follow-Up Respondent Sample,
Reading Apprenticeship Schools

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Race/ethnicity (%)				
Hispanic	31.5	31.8	-0.3	0.885
Black, non-Hispanic	43.1	43.8	-0.6	0.787
White, non-Hispanic	18.5	18.7	-0.2	0.916
Other	6.9	5.8	1.1	0.447
Gender (%)				
Male	50.0	51.7	-1.7	0.569
Female	50.0	48.3	1.7	0.569
Average age (years)	14.7	14.7	0.0	0.253
Overage for grade ^a (%)	27.0	25.2	1.8	0.475
Language other than English spoken at home (%)	45.0	45.1	0.0	0.991
Language spoken at home missing (%)	7.1	7.7	-0.5	0.701
Mother's education level (%)				
Did not finish high school	17.9	16.4	1.6	0.488
High school diploma or GED certificate	25.2	23.6	1.6	0.527
Completed some postsecondary education	27.8	30.1	-2.3	0.395
Don't know	21.3	21.7	-0.4	0.855
Missing	7.7	8.2	-0.5	0.740
Father's education level (%)				
Did not finish high school	16.8	17.1	-0.3	0.880
High school diploma or GED certificate	21.6	23.3	-1.7	0.491
Completed some postsecondary education	17.5	19.5	-2.0	0.395
Don't know	35.7	30.1	5.6 *	0.050
Missing	8.5	10.0	-1.6	0.331
GRADE reading comprehension ^b				
Average standard score	86.0	86.1	0.0	0.878
<i>Corresponding grade equivalent</i>	5.2	5.2		
<i>Corresponding percentile</i>	17	17		
6.0 - 7.0 grade equivalent (%)	36.4	35.5	0.9	0.743
5.0 - 5.9 grade equivalent (%)	29.0	28.0	1.0	0.712
4.0 - 4.9 grade equivalent (%)	34.5	36.5	-2.0	0.495
Sample size	686	454		

(continued)

Table 2.5 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year.

The differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is the ERO group value minus the difference.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

^bThe 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 A). No statistical tests or arithmetic operations were performed on these reference points.

assignment process. The non-ERO group serves as a benchmark or counterfactual for how students in the ERO group would have performed if they had not had access to the programs. Therefore, the impacts (differences in outcomes between the ERO and the non-ERO groups) represent the effect that the ERO programs had students' reading achievement and other outcomes over and above what the students would have achieved had they stayed in their regularly scheduled elective class.

This section of the chapter discusses several technical issues that lie at the heart of the evaluation's capacity to produce valid and reliable estimates of the literacy interventions' impacts on student reading achievement and other outcomes. It first reviews the study's sample sizes and the implications for statistical power (that is, the precision with which the analysis can measure program impacts). The section then reviews the estimation model being used to generate impacts and finally discusses the standards used for indicating statistical significance (that is, the confidence one may have that the impact estimates are not zero).

Sample Sizes and Statistical Power

To ensure that the ERO impact evaluation could produce valid and reliable findings, several design features were put in place to enable the study to measure program effects (if they exist) that are large enough to be both meaningful in students' lives and relevant to policy debates about the efficacy of supplemental literacy interventions.²⁷ The number of schools and the number of student sample members are crucial factors that determine the degree to which the impacts on student achievement and other outcomes can be estimated with enough precision to

²⁷Appendix C provides a more detailed assessment of the statistical power of the ERO study's impact design and discusses the role of other design features and assumptions, including the use of pre-random assignment characteristics to improve precision and assumptions about fixed versus random effects.

The Enhanced Reading Opportunities Study

Table 2.6

Characteristics of Students in Cohort 1
Follow-Up Respondent Sample,
Xtreme Reading Schools

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Race/ethnicity (%)				
Hispanic	33.8	34.2	-0.4	0.838
Black, non-Hispanic	42.7	43.5	-0.8	0.686
White, non-Hispanic	18.0	15.9	2.2	0.239
Other	5.5	6.5	-1.0	0.463
Gender (%)				
Male	50.1	51.0	-0.9	0.762
Female	49.9	49.0	0.9	0.762
Average age (years)	14.8	14.7	0.0	0.244
Overage for grade ^a (%)	29.1	25.2	3.9	0.104
Language other than English spoken at home (%)	49.0	46.8	2.3	0.365
Language spoken at home missing (%)	6.4	6.7	-0.4	0.749
Mother's education level (%)				
Did not finish high school	16.1	17.1	-1.0	0.627
High school diploma or GED certificate	25.8	25.6	0.1	0.959
Completed some postsecondary education	30.2	32.3	-2.1	0.395
Don't know	20.6	17.8	2.8	0.197
Missing	7.3	7.2	0.2	0.893
Father's education level (%)				
Did not finish high school	16.6	16.7	-0.1	0.968
High school diploma or GED certificate	24.2	21.2	3.0	0.206
Completed some postsecondary education	18.8	24.4	-5.6 *	0.013
Don't know	32.3	29.7	2.6	0.322
Missing	8.0	7.9	0.1	0.930
GRADE reading comprehension ^b				
Average standard score	85.7	86.3	-0.5	0.058
<i>Corresponding grade equivalent</i>	5.1	5.2		
<i>Corresponding percentile</i>	16	17		
6.0 - 7.0 grade equivalent (%)	32.5	38.2	-5.6 *	0.036
5.0 - 5.9 grade equivalent (%)	29.6	25.1	4.5	0.068
4.0 - 4.9 grade equivalent (%)	37.8	36.8	1.1	0.690
Sample size	722	551		

(continued)

Table 2.6 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year.

The differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is the ERO group value minus the difference.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

^bThe 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 A). No statistical tests or arithmetic operations were performed on these reference points.

reject with confidence the hypothesis that the program had no effect. In general, larger sample sizes provide more precise impact estimates.

An important goal for the design of the ERO study was to ensure that the sample sizes would be sufficient to allow for estimates of even “small” impacts on reading test scores and other outcomes both overall and for each of the supplemental literacy programs separately.²⁸ As discussed above, there are a total of 2,413 students in the Cohort 1 follow-up respondent sample for the impact analysis presented in this report. This includes 1,140 students from the 17 high schools using the Reading Apprenticeship program and 1,273 students from the 17 high schools using the Xtreme Reading program.

The overall study sample is equipped to detect impacts as small as 0.06 standard deviation units (referred to as “effect sizes”).²⁹ These pooled impact estimates provide insight into the impact of the family of interventions that share characteristics with Reading Apprenticeship and Xtreme Reading. The samples for each of the two supplemental reading programs are equipped to detect impacts as small as approximately 0.10 effect size.

²⁸There are no universally agreed-upon standards for what constitutes “small” versus “large” impacts. Some attempts have been made to examine the range of effects that have been found across a wide array of evaluations and to divide this range into segments that reflect the higher, middle, and lower categories of effects (see Lipsey, 1990). More recent work has begun to examine actual year-to-year rates of growth on a variety of achievement measures for students in a range of school districts and with a variety of background characteristics (see Bloom, Hill, Black, and Lipsey, 2006). These analyses provide additional background for interpreting the impact of interventions like those in the ERO study within the context of the expected growth in student outcomes nationally and under similar conditions.

²⁹The actual precision of estimated impacts may differ somewhat from those calculated for the statistical power analyses presented in Appendix C. These differences are due to such factors as actual variation in sample sizes, random assignment ratios, pretest scores, and outcomes levels across sites.

Statistical Model for Estimating Impacts

The ERO study impact analysis uses the following statistical model to estimate impacts on both reading achievement and reading behaviors:

$$Y_i = \sum_n \gamma_{0n} S_{ni} + \gamma_1 Y_{-i} + \sum_S \gamma_{2s} X_{si} + \beta_0 T_i + \varepsilon \quad (1)$$

Where:

Y_i = reading achievement or reading behaviors outcome for student i

$\sum_n S_{ni}$ = school dummy variable, one if student i is in school n and zero otherwise

Y_{-i} = the GRADE reading comprehension test score for student i before random assignment

$\sum_s X_{si}$ = other pre-random assignment characteristics for student i

T_i = one if student i is assigned to the ERO group and zero otherwise

ε_i = student-level random error term

In this model, β_0 represents the estimated impact of the ERO programs on the outcome of interest (Y_i). β_0 is a fixed-effect impact estimate that addresses the question: What is the impact of the ERO programs for the average student in the follow-up respondent sample? This approach is taken because this study most closely reflects an efficacy study of the effects of a new supplemental literacy intervention under relatively controlled conditions. Also, the sites and students were not selected to be a random sample of a larger population of sites. Instead, sites were selected purposively through the OVAE special SLC grant competition using specific criteria that differentiated these schools and districts from others that were not awarded a grant. In short, the impact estimates are not statistically generalizable to a larger population of districts, high schools, or students. As discussed above, however, on average, the participating schools share characteristics of other low-performing urban high schools across the country.

Equation 1 includes indicator variables for each of the participating high schools. These covariates capture a central feature of the study design in which random assignment was conducted within each of the participating high schools. These covariates are included to account for variation in the mean value of the dependent variable across the participating high schools.

Equation 1 also includes a covariate for each student’s GRADE reading comprehension test score at baseline and a covariate indicating whether the student is overage for grade (and likely to have been retained in a prior grade). These covariates are included to improve the precision of the impact estimates.

Statistical Significance

Equation 1 is estimated using ordinary least squares (OLS) regression, and a two-tailed t-test is used to assess the statistical significance of the impact estimate (β_0). *Statistical significance* is a measure of the degree of certainty one may have that some non-zero impact actually occurred. If an impact estimate is statistically significant, then one may conclude with some confidence that the program really had an effect on the outcome being assessed. If an impact estimate is not statistically significant, then the non-zero estimate is more likely to be a product of chance. For the purposes of this report, statistical significance is indicated in the tables by an asterisk (*) when the p-value of the impact estimate is less than or equal to 5 percent.

When making judgments about statistical significance, it is also important to recognize potential problems associated with conducting multiple hypothesis tests. Specifically, the analysis should avoid concluding that an impact estimate is statistically significant when in fact, there is no true impact (that is, relying on false positive results.) Likewise the analysis should not be so conservative with respect to producing false positives that it unduly increases the likelihood of missing true impacts when they exist (that is, relying on false negative results).

The statistical significance of the impact estimates presented in this report should be interpreted in light of two sets of safeguards aimed at attenuating the risk of drawing inappropriate conclusions about program effectiveness on the basis of ancillary hypothesis tests or statistically significant results that may occur by chance.³⁰ The first safeguard was to confine the analysis to a parsimonious list of outcome measures and subgroups. The shorter this list, the fewer the number of hypothesis tests and, thus, the less exposed the analysis will be to “spurious statistical significance” as a result of having tested multiple hypotheses. The primary evidence of overall ERO program effectiveness for this report will be reflected by estimates of program impacts on reading comprehension test scores (expressed in standard score values) for the full study sample and for each of the two ERO programs being evaluated. Vocabulary knowledge and student reading behaviors, while targets of the interventions and important to students’ literacy development, are considered secondary indicators of program effectiveness. Similarly, subgroups of students and subgroups of schools provide useful information about the relative impact of supplemental literacy programs, but they too are considered secondary indicators of effectiveness in this report.

³⁰See Appendix E for a more detailed discussion of the approach used to address the risks associated with multiple hypothesis tests.

The second safeguard uses composite statistical tests to “qualify” or call into question multiple hypothesis tests that are statistically significant individually but that may be due to chance in the context of mixed results.³¹ In general, these qualifying statistical tests estimate impacts on composite indices that encompass all the measures in a given domain or estimate the overall variation in impacts across subgroups.³² If the results of these tests are not statistically significant, this indicates that the statistical significance of the associated individual impact estimates *may have* occurred by chance. In these cases, the discussion of the impacts includes cautions or qualifiers about the robustness of the individual findings.

Finally, statistical significance does not directly indicate the magnitude or importance of an impact estimate — only the probability that an impact may have occurred by chance. Some statistically significant impacts may not be seen as policy relevant or as justifying the additional costs and effort to operate the programs under study. As a result, it is sometimes useful to frame the impact estimates in terms of other benchmarks and contexts, such as improvements found for related constructs or interventions, cost-effectiveness indicators, achievement gaps, or performance standards, which can help policy makers, practitioners and researchers gauge the importance or relevance of the findings. By the same token, lack of statistical significance for an impact estimate does not mean that the impact being estimated equals zero. It only means that the estimate cannot be distinguished from zero reliably. This can be due to the small magnitude of the impact estimate, the limited statistical power of the study, or some combination of both.

³¹Measurement of overall effects has its roots in the literature on meta-analysis (see O’Brien, 1984; Logan and Tamhane, 2003; and Hedges and Olkin, 1985). For a discussion of qualifying statistical tests to account for the risk of Type I error, see Duflo, Glennerster, and Kremer (2007). Other applications of these approaches are discussed in Kling and Liebman (2004) and Kling, Liebman, and Katz (2007).

³²See Appendix E for a more detailed description of the method used to conduct these qualifying statistical tests. Appendix E also includes tables with the results of these tests.

Chapter 3

Implementing the Supplemental Literacy Programs

This chapter describes the two supplemental literacy programs that are being used in the high schools participating in the Enhanced Reading Opportunities (ERO) study and assesses the fidelity of their implementation during the first year of the study. The chapter's first section provides an overview of the process used to select the programs at the start of the study and then describes the programs' core elements as presented in the proposals submitted by their developers and in other literature and materials associated with the programs. The second major section of the chapter presents the background characteristics of the teachers who elected to teach the ERO classes and describes the training activities and technical support they received to prepare them for this work. The third section of the chapter discusses findings on the fidelity with which each of the supplemental literacy programs was implemented in the participating high schools. The chapter concludes with a discussion of factors affecting the first year of implementation and how the second year of implementation has been different.

There are several key points made in this chapter:

- The two programs evaluated were selected by an independent national panel of adolescent literacy experts from among 17 proposals through a competitive process.
- Both programs focus on establishing a positive learning environment in the classroom to facilitate the delivery of instruction in reading comprehension processes and strategies. The comprehension instruction seeks to make explicit the processes used by capable readers, teaching less proficient students to pay attention to *how* they read so that they can improve their comprehension of *what* they read.
- Teachers self-selected to teach the ERO programs and were approved by the schools, districts, and ED. They held a high school teaching license or certificate and had an average of over 11 years of teaching experience. Three of the 34 starting teachers discontinued their involvement in the study before the end of the school year, and their replacements were trained and provided with coaching as they took over the ERO classes.
- The implementation of the ERO programs in 16 of the 34 participating high schools was deemed to be “well aligned” with the respective program models in the first year. Eight of the schools were found to have achieved a level of

implementation “moderately aligned” with both the classroom learning environments and the reading comprehension instruction practices specified by the developers. Implementation of the ERO programs in the remaining 10 high schools was found to be problematic, and either the classroom learning environments and/or the comprehension instruction practices were deemed to be “poorly aligned” with the models specified by the developers.

Characteristics of the Supplemental Literacy Programs: Reading Apprenticeship Academic Literacy and Xtreme Reading

The supplemental literacy programs were selected through a competitive proposal process that was managed by the study team and guided by a panel of seven nationally known experts in adolescent literacy research and program development. A request for proposals (RFP) was advertised in a wide range of education publications and was disseminated to over 40 organizations that develop and implement high school curricula.¹ The RFP specified that prospective supplemental literacy programs must be research-based, high-quality programs that provide instruction in the areas that experts increasingly agree are necessary for effective adolescent literacy instruction, as outlined in *Reading Next*, but that were not yet rigorously tested.² The prospective programs were to have been developed already (that is, not be new programs) and to be ready for systematic use in multiple schools and districts.

Seventeen proposals were submitted in response to the RFP. After a review of the research base presented in the proposals for each program, the proposals were rated by the panel of adolescent literacy experts. The developers of four of the proposed programs were invited to give oral presentations before the panel, staff from the U.S. Department of Education (ED), and the ERO study team. Based on the presentations and subsequent discussion, the panelists recommended and ED accepted two programs for inclusion in the study: WestEd’s Reading Apprenticeship Academic Literacy and the University of Kansas Center for Research on Learning’s (KU-CRL) Xtreme Reading.

Overall Goals and Approach

The overarching goal of both Reading Apprenticeship and Xtreme Reading is to help students adopt the strategies and routines used by proficient readers, improve their comprehension skills, and motivate them to read more and enjoy reading. Both programs emphasize the importance of establishing a specific type of learning environment in the classroom that is con-

¹American Institutes for Research (2004).

²Biancarosa and Snow (2004).

ducive to the effective delivery of the core instructional strategies by the teacher and to facilitate student and teacher interactions around the reading skills that are being taught and practiced. They both use a “cognitive apprenticeship” approach to instruction in which the teacher initially takes the lead in modeling the strategies that proficient readers use and then gradually increases the responsibility of the students to demonstrate and apply these strategies. The teachers seek to make explicit *how* proficient readers read, and they support their students in recognizing and using the strategies or methods used by stronger readers. That is, both programs focus students’ attention on *how* they read (a metacognitive process) to help the students better understand *what* they read (understanding content). Also, both programs integrate direct, whole-group instruction with small-group and individualized instruction.³

Key Components

The key components of Reading Apprenticeship and Xtreme Reading are discussed categorically below. This discussion is based on information provided by the two program developers. Table 3.1 also presents these components by category. These components are the specific aspects of the programs’ instructional approaches that the developers expect to improve the literacy skills of high school students.⁴

Developer’s Implementation Philosophy

In implementing Reading Apprenticeship, teachers are 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 they expect to be motivating to their classes. Teachers have flexibility in how they include various aspects of the Reading Apprenticeship curriculum in their day-to-day teaching activities, but have been trained to do so such that they maintain the overarching spirit, themes, and goals of the program in their instruction.

Xtreme Reading was developed with the philosophy that the presentation of instructional material — particularly the order and manner in which the material is presented — is of critical import to the students’ understanding of it, and as such teachers are trained to deliver course content and materials in a precise, organized, and systematic fashion designed by the

³Additional information about the Reading Apprenticeship Academic Literacy course is available on the Internet at <http://www.wested.org/cs/we/view/serv/111>; information about the Xtreme Reading course is available at <http://www.xtremereading.org/>. Furthermore, the descriptive material about the program-specific observation rating scales in Appendix D provides more information specific to each program.

⁴The proposals submitted by the two developers, WestEd (2004) and University of Kansas (2004), contain information about the key components of their programs. These proposals are unpublished and cannot be released based on the rules of the competition through which the programs were selected.

The Enhanced Reading Opportunities Study

Table 3.1

Key Components of the ERO Programs

	WestEd/Reading Apprenticeship	KU-CRL/Xtreme Reading
Developer’s Implementation Philosophy	“Flexible fidelity” guided by the instructional and behavioral/social needs of the students	Prescribed daily lesson plans and time limits on classroom activities
Role of Teacher	Instructor as “master reader,” apprenticing students in various literacy competency areas and drawing on variety of materials	Instructor explicitly teaches seven reading strategies using a prescriptive eight-stage instructional approach with step-by-step instructional materials
Curriculum Design	<p><u>Learning Environment</u> Establish “social reading community” early in program</p> <p><u>Comprehension Instruction</u> Five curricular strands of classroom instruction:</p> <ol style="list-style-type: none"> 1. Metacognitive Conversation 2. Silent Sustained Reading 3. Language Study 4. Content/Theme 5. Writing 	<p><u>Learning Environment</u> Focus at beginning of course on teaching social and behavioral skills and strategies aimed to develop a productive and positive classroom learning environment</p> <p><u>Comprehension Instruction</u> Focus of rest of course on developing literacy skills through seven learning strategies:</p> <ol style="list-style-type: none"> 1. LINC’S Vocabulary Routine 2. Word Mapping 3. Word Identification 4. Self-Questioning 5. Visual Imagery 6. Paraphrasing 7. Inferencing
Teaching Strategies	Instructors usually use one or two of the following routines during class period: <ol style="list-style-type: none"> 1. Think aloud 2. Talking to the text 3. Metacognitive logs/journals 4. Preambles (daily warm-ups) 	Each strategy is taught using a prescribed eight-stage instructional methodology: <ol style="list-style-type: none"> 1. Describe 2. Model 3. Verbal practice 4. Guided practice 5. Paired practice 6. Independent practice 7. Differentiated instruction 8. Integration and generalization
Program Type	Supplemental course, like an elective	Supplemental course, like an elective
Duration	One school year	One school year

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. However, there are opportunities in the Xtreme Reading instructional program for teachers to use responsive instructional practices to adapt and adjust to student needs that arise as they move through the highly structured curriculum.

Role of Teacher

Both Reading Apprenticeship and Xtreme Reading are grounded in the principle of a cognitive apprenticeship. That is, the teacher assumes the role of reading expert whose task is to share expertise in explicit ways with the students and then to support their development of those skills and nurture their increased independence in using them. The process is one that starts off as teacher-centered and gradually transitions to being student-centered. In Reading Apprenticeship — where the teacher is considered the “master reader” for the students, who are the “reading apprentices” — the transition is facilitated through the teacher’s integration of the four dimensions of classroom life (personal, social, cognitive, and knowledge-building; described below), which he or she links together through ongoing metacognitive conversations (thinking internally and talking externally about reading processes).

For the Xtreme Reading teacher, this transitional process is a specific eight-stage instructional model through which seven specific literacy strategies are taught. In Xtreme Reading classes, the expectation is that the learning of each strategy begins with specific teacher-directed instruction and that control is relinquished to students incrementally as they progress through the stages. By the eighth stage, students are working independently and have an understanding of the application of the strategy outside the Xtreme Reading classroom.

Curriculum Design and Teaching Strategies

As discussed above, the two programs are attentive to both the learning environment in the classroom and the nature of the literacy instruction, particularly around reading comprehension. The curriculum design and the teaching strategies of the two ERO programs reflect these two priorities. Table 3.1 provides an overview of the key elements of each ERO program. The developers’ curriculum designs both highlight the equal importance of creating a conducive classroom learning environment and focusing instruction on strategies that promote reading comprehension skills and proficiency.

The core of the Reading Apprenticeship program is the integration of four dimensions: social, personal, cognitive, and knowledge-building. The social and personal dimensions reflect the attention of the program to the *learning environment* for the class. The social dimension refers to adolescents’ interests in peer interaction and in larger social, political, and cultural issues. The personal component addresses students’ own goals for reading and for reading improve-

ment. These aspects of the program are combined in the establishment of a social reading community, a classroom environment that allows for the respectful, open exchange of ideas considered essential for the program to have effective comprehension instruction.

The cognitive and knowledge-building dimensions are the *instructional components* of the Reading Apprenticeship program. They address students' needs to increase both their repertoire of comprehension strategies and their background knowledge, expanding their knowledge base through reading, and providing knowledge about aspects of strong reading such as word construction, vocabulary, text structure, or figurative language. The instructional components are delivered across the following three major thematic units during the school year: "Who Am I as a Reader?" "Reading History," and "Reading Science and Technology." Within each unit, the teacher incorporates the five key curricular strands of the program:

- **Metacognitive conversations.** The students and the teacher think and talk about the thinking processes that are engaged when reading.
- **Silent sustained reading.** The student reads a book of his or her choice for 20 to 25 minutes at least twice a week to build reading fluency, comprehension, motivation, and stamina.
- **Language study.** The teacher and the students routinely practice strategies and learn skills at the word, sentence, and text levels to enhance language development.
- **Content/theme.** The teacher uses the majority of instructional time to address one of the three thematic units of the curriculum so that students are able to apply what they are learning in the classroom to their other classrooms and relate what they are learning to contexts other than Reading Apprenticeship.
- **Writing.** The teacher provides opportunities for the students to write and provides new knowledge of writing processes and strategies as needed.

The curriculum strands are taught and reinforced through the use of four teaching strategies: *think alouds*, *talking to the text*, *metacognitive logs*, and *daily preambles*. These strategies offer teachers and students opportunities to interact around what they are reading and how they are reading.

The Xtreme Reading program also emphasizes creating a positive *learning environment* in the classroom. The program aims to create a structured classroom climate with explicit social and behavioral expectations and regular routines for both students and teachers. The main tenet of classroom management is time-on-task behavior; this is essential to successful implementa-

tion of the instructional sequence. Student motivation and engagement are encouraged through several activities that help students set short- and long-term goals for their learning and through the availability and sharing of high-interest novels about students who have overcome academic obstacles. Teachers seek to help students to set real purposes for learning and to link their learning to personal goals.

The program's *literacy instruction* involves both a systematic component (driven by the curriculum) and a responsive component (driven by student needs). The systematic component involves teaching seven reading strategies following lesson plans provided by the developer that map out daily instruction. Two strategies focus explicitly on vocabulary: LINC and Word Mapping. Five strategies focus more directly on comprehension: Word Identification, Self-Questioning, Visual Imagery, Paraphrasing, and Inferencing. Each strategy is taught using an eight-stage model that starts off being highly teacher-centered (the teacher describes and models the strategy in the first two stages), to being shared work between the teacher and the students (verbal and guided practice), to being more and more the responsibility of the students (paired practice between students and independent student practice). The seventh stage is differentiated instruction, allowing those struggling with the strategy to receive additional support and those who have been successful learning the strategy more and varied opportunities for practice. The eighth stage, integration and generalization, involves students' taking the strategy beyond the Xtreme Reading classroom and materials and applying it to reading in other classes. The responsive instruction component focuses on assessing and addressing individual student needs as they arise. The responsive instruction component represents where flexibility enters into Xtreme Reading instruction.

Both ERO programs were developed from preexisting programs prior to implementation in the ERO study. The program developers adapted their already existing curricula to create programs that would be supplemental, yearlong reading classes. The Reading Apprenticeship Academic Literacy curriculum combined elements of two WestEd programs, Reading Apprenticeship and Academic Literacy. These programs had been the focus of most of the work within WestEd's Strategic Literacy Instruction initiative. Instruction in Reading Apprenticeship helps students identify weaknesses in their reading skills and improve them through mastering and then consciously applying advanced reading strategies. Academic Literacy is usually woven into content-area instruction so that students learn to apply subject-specific skills and strategies in areas such as science and social studies. The curriculum used in this study offered instruction in strategic reading within three themed units, two of which emphasized content-area reading. The Xtreme Reading curriculum combined the components of the Strategic Instruction Model (SIM) for reading improvement that has been developed, studied, and refined at the University of Kansas Center for Research on Learning for close to 30 years. SIM content consists of six specific reading processes, such as vocabulary identification and strategies for making inferences from the text. Previous implementation of SIM had followed the eight-stage instructional

model used in Xtreme Reading but had not combined the six reading strategies into a full-year curriculum for use in self-contained intervention classes. Further, two versions of this curriculum were developed to accommodate both 45- and 90-minute instructional blocks.

The ERO Teachers and Their Preparation for the ERO Programs

Teachers play a key role in both programs selected for the study. The study sought to have experienced, core-content-area teachers implement the programs and to provide adequate training and support for them. The teachers were nominated by their schools on the grant applications submitted to the Office of Vocational and Adult Education (OVAE) at ED. Additionally, participating districts and schools committed to make these teachers available for professional development activities prior to the start of the school year and on an ongoing basis during the year.

Teacher Characteristics

The Request for Proposals from OVAE to which school districts responded in their application for grant funding and participation in this study specified that teachers selected to teach the ERO classes at each high school should have at least two years of experience and be certified core-content-area teachers — specifically, English or social studies teachers — and not necessarily reading specialists. The project sought to target content-area teachers rather than reading teachers to teach the classes in order to enhance the replicability of the interventions if they proved to be effective. First, the study sought to demonstrate that if content-area teachers could be trained to deliver a literacy program, schools and districts that later chose to pursue this type of intervention may have a more realistic chance to identify staff to teach it without being restricted to reading specialists. Second, one of the goals of both interventions is transference — helping students use the literacy skills that they develop in their content-area classes. Thus, it was hoped that involving content-area teachers would help facilitate this.

Table 3.2 provides a list of background characteristics for the teachers in each of the two ERO programs.⁵ The average number of years of previous experience for ERO teachers was 11.2 years, although prior teaching experience ranged from student teaching to over 30 years as a regular classroom teacher. Almost three-quarters (73.5 percent) of the teachers had graduate-level degrees, and almost all (97.1 percent) held high school-level certification. The majority of the teachers (76.5 percent) were certified in English/language arts, with nearly 18

⁵Information in Table 3.2 is drawn from the survey that teachers completed at the beginning of the ERO training or at the beginning of their tenure as an ERO teacher. The information in the table reflects the characteristics of the teacher who spent the longest period of time as the ERO teacher in each participating school. Three of the teachers who began the 2005-2006 school year teaching the ERO students left that position before the end of the school year.

The Enhanced Reading Opportunities Study

Table 3.2

Background Characteristics of ERO Teachers

Characteristic	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Race/ethnicity (%)			
Black	20.6	23.5	17.7
White	67.7	64.7	70.6
Other	11.8	11.8	11.8
Gender (%)			
Male	23.5	11.8	35.3
Female	76.5	88.2	64.7
Total time teaching (years) ^a	11.2	9.0	13.5
Total time teaching at current school (years) ^b	4.8	4.7	4.9
Total time teaching at current level (years) ^a	7.1	5.7	8.6
Total time teaching English/language arts or social studies (years) ^a	10.4	8.4	12.7
Master's degree or higher (%)	73.5	70.6	76.5
Holds high school-level teaching certification (%)	97.1	100.0	94.1
Subject matter certification (%)			
Certified in English/language arts	76.5	70.6	82.4
Certified in social studies	17.7	23.5	11.8
Certified in other subject	5.9	5.9	5.9
Number of professional development workshops attended in the last two years ^a	3.8	4.2	3.3
Number of hours spent in professional development workshops during the last two years ^b	45.4	40.9	50.4
Taught the ERO class for the full school year (%)	91.2	100.0	82.4
Sample size	34	17	17

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study baseline teacher survey.

NOTES: For three schools, the original teacher was replaced during the school year. The table includes the teacher who spent the most time teaching the ERO program.

Rounding may cause slight discrepancies in calculating sums and differences.

^aMissing data: One to two teachers did not respond.

^bMissing data: Four to five teachers did not respond.

percent holding social studies certification and 6 percent holding certification in some other area. Teachers reported attending an average of 45.4 hours of professional development in the two years prior to the beginning of the ERO program.⁶

Training and Technical Assistance

Training and technical assistance were delivered to the ERO teachers in the following ways: Reading Apprenticeship teachers attended one 5-day summer training institute as well as two 2-day booster training sessions during the 2005-2006 school year. They also received ongoing support through three 2-day coaching visits during the year and access to a special online listserv that was set up for the project. Xtreme Reading teachers attended one 5-day summer training and one 2-day booster training during the year. They also received three 2-day on-site coaching visits. District coordinators were asked to attend the trainings to familiarize them with the programs in case they had to provide technical assistance or other support to ERO teachers. Table 3.3 summarizes the activities provided by each of the developers for the 2005-2006 school year.

The Enhanced Reading Opportunities Study

Table 3.3

Training and Technical Assistance Provided During the 2005-2006 School Year, by ERO Program

	Summer Training	School-Year Booster Training	Additional Supports
Reading Apprenticeship	One 5-day training (August)	Two 2-day trainings (November; February)	Three 2-day on-site coaching visits Weekly e-mail and phone calls Listserv
Xtreme Reading	One 5-day training (August)	One 2-day training (January)	Three 2-day on-site coaching visits Weekly e-mail and phone calls Additional technical assistance for replacement teachers

⁶Differences between teachers in each ERO program were not tested for statistical significance. There is one ERO teacher per school, which means that teacher characteristics are also school characteristics. As discussed later in the chapter, the impact analysis accounts for differences across school characteristics (and, thus, across teachers) by including regression covariates for each school.

Summer Trainings

The summer teacher training institutes for both programs were conducted in August 2005. The Reading Apprenticeship Academic Literacy training was conducted by the program developer, experienced Reading Apprenticeship teachers, and the two coaches who would work with the ERO teachers throughout the year. The Xtreme Reading training was conducted by the program developers, research staff from the University of Kansas Center for Research on Learning, and the coaches who would work with the teachers throughout the year. Each of the trainings provided the teachers with an introduction to the program as a whole but also included time focused on the curricular units to be taught during the first quarter of the course. Training methods across both summer institutes included modeling, discussion, and formal presentations as well as large-group and small-group activities. Teachers also had time to meet with the coaches with whom they would be working during the year. Fifteen of the 17 Reading Apprenticeship teachers attended the summer training. The other two attended national Reading Apprenticeship workshops before they began teaching the course.⁷ All of the Xtreme Reading teachers attended the summer training session.

Items on surveys administered to the teachers at the conclusion of the summer training probed the teachers' perceptions of their preparedness for teaching the ERO classes and their sense of the challenge they faced in implementing the programs. Thirty-three of the 34 ERO teachers (one teacher did not respond to the item) agreed or strongly agreed with the statement "I will be able to present this program confidently to students with the help of the manuals, other materials, and support of the professional developers." Additionally, 29 of the ERO teachers (15 of the 17 Reading Apprenticeship teachers and 14 of the 17 Xtreme Reading teachers) disagreed or strongly disagreed with the statement that the "[Reading Apprenticeship and Xtreme Reading]-recommended strategies and activities seem difficult to implement." Of the other five ERO teachers, two did not respond to the item and three agreed or strongly agreed that it would be difficult to implement the programs' strategies and activities.

Booster Trainings

The booster trainings during the school year (two for Reading Apprenticeship and one for Xtreme Reading) were conducted in a similar format to the summer training institutes and were two days each in duration. The program developers introduced the teachers to the curricular units coming up next in the programs as well as to the computer-based components of the

⁷The Reading Apprenticeship Academic Literacy course being implemented in the ERO Study is an adaptation of the preexisting Reading Apprenticeship program on which the national workshops were focused. While at the national workshops, these two ERO teachers received additional training that addressed aspects of Reading Apprenticeship that are specific to the ERO Study.

courses. Each of the trainings also provided time for the teachers to meet with their coaches and opportunities for the teachers and developers to discuss any issues with the implementation of the program that had come up during the first part of the year. All 17 of the Reading Apprenticeship teachers attended both booster training sessions. Sixteen of the Xtreme Reading teachers attended the booster training session in person, and one teacher participated by telephone.

Ongoing Technical Assistance

Both programs provided on-site coaching and electronic communication among teachers and their coaches. Reading Apprenticeship also made a listserv available to teachers. The Reading Apprenticeship and Xtreme Reading coaches made three 2-day visits to each of the teachers, during which they observed classes, modeled instruction, and in some cases co-taught lessons, in addition to working through issues that each teacher was experiencing. In the three instances of teacher turnover, coaches provided additional technical assistance to the replacement teachers.

Implementation Fidelity

This section of the chapter examines the fidelity with which the two supplemental literacy programs — Reading Apprenticeship Academic Literacy and Xtreme Reading — were implemented. In particular, it defines the method by which composite measures of implementation fidelity were computed for each school, based on classroom observations conducted by study team members during site visits in the second semester of the first year of implementation. 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 approximated the intended learning environments and instructional practices that were specified by the model developers.

Overall ratings of the implementation fidelity of the ERO programs at each school provide a context for interpreting the study’s impact findings and offer information to policymakers and practitioners about factors they may wish to consider if establishing these programs or ones like them in high schools.

Data Sources and Measures

As noted in Chapter 2, the analysis of ERO program implementation fidelity in the first year of the study is based on field research visits to each of the 34 high schools during the second semester of the 2005-2006 school year.⁸ The classroom observation protocols used in the site visits provided a structured process for observers to rate characteristics of the ERO class-

⁸Appendix D provides more detailed description of the site visits.

room learning environments and the ERO teachers' instructional strategies. The instrument included ratings for six characteristics (referred to as "constructs" from here forward) that are common to both programs and ratings for seven program-specific constructs. The analysis of the classroom observation ratings sought to capture the implementation fidelity of two key overarching dimensions of both programs: the classroom learning environment and the instructional strategies that focused on reading comprehension. A composite measure of implementation fidelity for each dimension was calculated from the average ratings for both general and program-specific constructs.

Table 3.4 provides a list of the constructs that were combined to create composite ratings for the learning environment and comprehension instruction dimensions, respectively, for the ERO programs in each high school. The *learning environment composite* was calculated as the average of ratings on two general constructs and ratings of one or two program-specific constructs for Reading Apprenticeship and Xtreme Reading, respectively. The *comprehension instruction composite* was calculated as the average of ratings on two general constructs and ratings of five program-specific constructs.⁹ The composite measures ranged from one to three and were rounded to the nearest tenth of a point.

Based on the composite ratings for each of the two program dimensions — learning environment and comprehension instruction — the implementation fidelity for each dimension was classified as "well aligned," "moderately aligned," or "poorly aligned" with the models specified by the program developers. The fidelity analysis focused on identifying schools where implementation of one or both of the two key program dimensions was especially problematic. This focus is particularly relevant to the first year of implementation, when the programs were new to the schools and the teachers and their lack of prior experience with the programs presented a more challenging implementation scenario. Thus, the definitions below for each level of implementation fidelity include not only information about average ratings but also the number of constructs rated in Category 1 — implementation that was poorly aligned with the expectations of the ERO programs.

Implementation fidelity for the learning environment or comprehension instruction dimensions was characterized as *well aligned* when the average rating across the relevant general and program-specific constructs was 2.0 or higher. That is, the school's ERO program was rated as moderately (a Category 2 rating) or well aligned (a Category 3 rating) with the program models on all or almost all of the constructs included in that dimension. As it turned out, the

⁹Note that, for Xtreme Reading, the program-specific component comprises two subcomponents: curriculum-driven or systematic instruction and needs-driven or responsive instruction. Appendix D provides a detailed description of the method used to average the ratings on individual constructs to create the composites for the two overarching program dimensions.

The Enhanced Reading Opportunities Study

Table 3.4

**Dimensions and Component Constructs of Implementation Fidelity,
by ERO Program**

Dimension	Component	Reading Apprenticeship	Xtreme Reading
Learning Environment	General Instructional Constructs	Classroom climate On-task participation	Classroom climate On-task participation
	Program-Specific Constructs	Social reading community	Classroom management Motivation and engagement
Comprehension Instruction	General Instructional Constructs	Comprehension Metacognition	Comprehension Metacognition
	Program-Specific Constructs	Metacognitive conversations Silent sustained reading Content/theme integration Writing Integration of curriculum strands	Curriculum-driven (systematic) instruction <ul style="list-style-type: none"> • Structured content • Research-based methodology • Connected, scaffolded, informed instruction Needs-driven (responsive) instruction <ul style="list-style-type: none"> • Student accommodations • Feedback to students

schools with fidelity rated as *well aligned* had no more than one construct for each implementation dimension rated in Category 1.

The key dimensions were designated as *moderately aligned* in terms of implementation fidelity if the average rating across the general and program-specific constructs used to create the relevant composite was within the range of 1.5 to 1.9. In these cases, the school's ERO program was observed to have some problems with implementation. In terms of learning environment, these schools had one construct rated in Category 1 (out of three or four constructs used to calculate the composite for Reading Apprenticeship or Xtreme Reading schools, respectively). On the comprehension instruction dimension, schools had three or fewer constructs rated in Category 1 (out of seven constructs used to calculate the composite score). These schools also met with some implementation success, with half or more of the constructs that make up the dimension being rated as moderately or well aligned with the program models.

The implementation fidelity of key program dimensions in a school was rated as *poorly aligned* when the average composite rating across the general and program-specific constructs fell below 1.5. This resulted when the school's ERO program was rated in Category 1 for half or more of the general or program-specific constructs that make up the dimension. These programs were the least representative of the activities and practices intended by the respective program developers.

The ratings and resulting categories indicate whether the programs reflected the characteristics of the classroom learning environments and instructional strategies intended by the developers. While it is reasonable to expect that higher fidelity programs could produce stronger impacts than programs where the fidelity was only a limited reflection of the intended model, other factors could intervene to make higher fidelity programs ineffective or to make limited or inadequate fidelity programs effective.

Findings

Table 3.5 provides a summary of the findings regarding implementation fidelity. The top two panels of the table provide a summary of the number of schools whose composite rating on the classroom learning environment and comprehension instruction dimensions fell into the well-aligned, moderately aligned, and poorly aligned categories of fidelity. The bottom panel of the table categorizes schools in terms of their overall implementation fidelity, based on their ratings across both implementation dimensions. The discussion that follows focuses first on each implementation dimension and then turns to overall fidelity, which accounts for the importance of the implementation of both dimensions to the ERO programs.

The Enhanced Reading Opportunities Study

Table 3.5

Number of ERO Classrooms Well, Moderately, or Poorly Aligned to Program Models on Each Implementation Dimension, by ERO Program

Implementation Dimension	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
<u>Learning environment</u>			
Well-aligned implementation (composite rating is 2.0 or higher)	26	14	12
Moderately aligned implementation (composite rating is 1.5-1.9)	4	2	2
Poorly aligned implementation (composite rating is less than 1.5)	4	1	3
<u>Comprehension instruction</u>			
Well-aligned implementation (composite rating is 2.0 or higher)	16	7	9
Moderately aligned implementation (composite rating is 1.5-1.9)	9	4	5
Poorly aligned implementation (composite rating is less than 1.5)	9	6	3
<u>Combined dimensions</u>			
Well-aligned implementation on both dimensions	16	7	9
Moderately aligned implementation on at least one dimension and moderately or well-aligned implementation on the other dimension	8	4	4
Poorly aligned implementation on at least one dimension	10	6	4
Sample size	34	17	17

SOURCES: MDRC and AIR calculations from classroom observation data.

NOTES: Implementation with a composite score of less than 1.5 for a given dimension was deemed to be at the beginning stages of development. The implementation for these dimensions was designated as poorly aligned with the program models.

Implementation with composite scores between 1.5 and 1.9 for a given dimension exhibited at least moderate development in some areas while being at the beginning stages of development in other areas. The implementation for these dimensions was designated as moderately aligned.

Implementation with scores of 2.0 or higher for a given dimension exhibited well-developed fidelity on several areas and at least moderate development in most other areas. The implementation for these dimensions was designated as well aligned.

Fidelity by Implementation Dimension

As described earlier in the chapter, the first curriculum unit for both Reading Apprenticeship and Xtreme Reading focuses on the learning environment in the classroom. This focus involves setting expectations for the organization of the classroom, for how students should interact with the teacher and with their peers, and for the daily and weekly schedules of classroom activities. These same expectations are reinforced in each of the subsequent curriculum units. Table 3.5 shows that the ERO programs in 26 of the 34 high schools reached a level of implementation that was well aligned with the program models in terms of classroom learning environment dimension. Four schools were rated as demonstrating moderate alignment on this dimension, and four schools were rated as demonstrating poor alignment.

Compared with aspects of the ERO programs focused on the classroom learning environment, comprehension instruction evolves differently over the course of the year and varies across curriculum units. Although instructional strategies that focus on metacognition and content are incorporated in all the curriculum units, teachers were learning new instructional features of each ERO program continuously throughout the first year of implementation. As a result, it took a year of work with the ERO program for teachers to be exposed to and use the full repertoire of comprehension instruction strategies. As shown in Table 3.5, implementation was rated as well aligned on the comprehension instruction dimension for the ERO programs in 16 of the schools. Nine schools demonstrated moderate alignment, and nine schools demonstrated poor alignment, on the comprehension instruction dimension.

Differences in Fidelity, by Implementation Dimension

The pattern of findings shown in Table 3.5 indicates that more schools reached a level of well-aligned implementation fidelity on the learning environment dimension (26 schools) than on the comprehension instruction dimension (16 schools). Two hypotheses offer explanations for this observed difference in the fidelity achieved by schools on these two dimensions. First, this difference may reflect how these programs evolve during their implementation. The continuous and mutually reinforcing way that the elements of the classroom learning environment dimension are situated in the curriculum presents ongoing opportunities for teachers to refine their implementation of this dimension's elements and reach alignment with the program model. The elements of comprehension instruction are revealed in a more step-by-step way, unit by unit as the year progresses. Thus, teachers do not have the same continuous opportunity to refine their implementation of each instructional element. A second hypothesis for the difference in fidelity achieved on the two implementation dimensions is the difference in teachers' experience with teaching reading as opposed to developing a positive classroom environment. The instructional aspects of the programs were new to the teachers, who came to the program

predominantly from core-content-area teaching and not reading or literacy. However, the principles behind the learning environment dimension of the program models reflect principles often advocated for classrooms across subject areas, such as respect between individuals and creating a safe space for sharing opinions and ideas. The program developers emphasize the importance of both program dimensions, but it is useful for policymakers and practitioners to know that, in the implementation of these programs or similar ones, different aspects of the programs may develop more quickly than others.

Rating the Overall Fidelity of ERO Program Implementation

The bottom panel of Table 3.5 clusters schools based on their levels of implementation fidelity across both the classroom learning environment and the comprehension instruction dimensions. Because the classroom learning environments and comprehension instruction activities were designed to be interdependent and mutually reinforcing, the implementation of the ERO program in a given school was deemed to be well aligned with the program model overall only if both of these dimensions were rated in this category. The ERO programs in 16 of the 34 schools were found to have reached this level of implementation on both the classroom learning environment and the comprehension instruction dimensions. These schools did not necessarily represent exemplary versions of the ERO program model being used, although some of them did. While there is variation among these schools, the assessment of their implementation fidelity revealed that all constructs or all but one construct across both implementation dimensions were rated as either moderately (Category 2) or well aligned (Category 3) with the program models. These 16 schools include seven Reading Apprenticeship schools and nine Xtreme Reading schools.

In eight of the 34 high schools, the implementation of the ERO program was rated as moderately aligned with the program model for at least one of the two key program dimensions. It should be noted that, for these schools, neither of the dimensions was rated as poorly aligned. In fact, the classroom learning environment was rated as well aligned for the ERO programs in six of these schools, while the comprehension instruction was found to have reached a level of moderate alignment. In the remaining two schools, both the classroom learning environment and the comprehension instruction were rated as moderately aligned in terms of fidelity to the program model. Thus, in these eight schools where the ERO programs were designated as having reached a level of moderate alignment overall, at least seven out of up to 11 constructs included in the composites were rated as being moderately or well aligned according to the criteria presented in the observation protocols. These eight schools include four Reading Apprenticeship schools and four Xtreme Reading schools.

Schools identified as having especially problematic program implementation were those schools whose average fidelity rating on either the classroom learning environment dimension or

the comprehension instruction dimension was classified as implementation poorly aligned to the program models. The bottom panel of Table 3.5 also shows that 10 of the 34 high schools were found to have encountered serious implementation problems on at least one of the two key program dimensions during the first year of the study. Three of these schools demonstrated poorly aligned implementation on both the learning environment and the comprehension instruction dimensions; six demonstrated poorly aligned implementation only on the comprehension instruction dimension; and one demonstrated poorly aligned implementation only on the classroom learning environment dimension. These 10 high schools that encountered serious implementation problems include six Reading Apprenticeship schools and four Xtreme Reading schools.

Summary and First-Year Implementation Challenges

Both of the ERO programs were complex and multidimensional interventions being implemented by teachers who had no prior formal experience with supplemental reading instruction for adolescents. Each of the program developers provided a five-day summer training institute prior to the start of the first year of the study. During the school year, teachers attended two 2-day booster training sessions, and coaches from the developer teams made a minimum of three coaching visits to each teacher.

In all, the ERO programs in 24 of the 34 schools were found to have reached a level of implementation at least moderately aligned with the program models. The ERO programs in 16 of these schools were found to have reached a level of implementation well aligned with the models, indicating that almost all of the key implementation components were moderately aligned or well aligned with the characteristics of the program models. The implementation of the ERO programs in the remaining 10 schools was found to be especially problematic, and these programs were deemed to be poorly aligned reflections of their intended models.

ERO implementation in the 2005-2006 school year occurred in the context of three challenges that were distinctive to the first year of the project:

- The delayed start of the ERO classes in all schools
- The delayed acquisition of some prescribed program materials and resources
- The newness of the programs to the schools and the ERO teachers

As is discussed in Chapter 4, ERO classes began an average of six weeks after the start of the school year, and 16 of the participating schools started their ERO programs during the eighth week of school or later. As a result, more than two months had elapsed between the summer training institute and the start of the ERO classes. This caused disruptions in students' class schedules, and teachers were left an average of less than seven and a half months to try to

cover curricula intended for a nine-month school year. In response to this shorter time line, the developers were able to make some adjustments to compact their curricula. Nonetheless, teachers also were not able to get through all of the curricular units.

Each ERO classroom was intended to have the following components: a library, a file cabinet, a flipchart, an overhead projector, two computers, and a printer/scanner. These resources were to be purchased by the school district, using funds from its SLC grant. The ERO study team visited each of the participating schools within approximately four weeks of the start of the ERO classes and found that one or more of these classroom components were missing in 23 of the 34 schools. They communicated these findings to the district coordinators, reminding them of the expectation that the grant funds would be used to provide these components.

Until districts were able to provide the components, teachers made accommodations by borrowing overhead projectors or file cabinets, for example, until there were provided permanently. The most commonly missing items were computers. In these cases, the ERO teachers were advised by the program developers — whose staff were also making visits to sites and were aware of which teachers were missing materials — to postpone using the software programs they provided until the second semester. By the second semester of the year, all supplies had been provided to 27 of the 34 schools. The study team continued to communicate with the other seven high schools and their districts to encourage them to obtain the rest of their supplies.

All ERO teachers were new to the program they were trying to implement. They were learning the Reading Apprenticeship or Xtreme Reading program while teaching it, adding to the challenge of achieving high-fidelity implementation. In addition, three of the 34 teachers who attended the summer training institutes left their ERO teaching position before the end of the academic year.¹⁰ The schools that lost teachers had to conduct a search for replacements who met the eligibility criteria for the project (holding a high school teaching certificate in social studies or English and having at least a year of teaching experience).¹¹ These teachers were then trained in the relevant ERO program.

Each of these challenges was addressed systematically in the second year of the study. ERO classes began within an average of approximately two weeks of the start of the school year and started on the first day of school at 18 of the 34 schools. All the required equipment and supplies were provided to each of the ERO classrooms. Twenty-seven of the 34 teachers of the ERO classes at the end of the first year of implementation returned to teach the program again

¹⁰One of the three departing teachers left after having participated in the summer training but before the ERO course had started. The other two teachers left approximately half way through the school year.

¹¹The study team worked with the U.S. Department of Education officials responsible for the grant administration and the evaluation and the grantees to identify suitable replacement teachers and to schedule them for training and coaching.

in the second year.¹² All of the continuing and replacement teachers remained with the programs throughout the second year of the study. Thus, the second report from the study will provide information about both the implementation and impact of the ERO programs under conditions of a timelier start-up, better-equipped classrooms, and more experienced teachers than existed in the first year of implementation. In fact, results from classroom observations in the fall of the second year — the first of two second-year site visits — indicate that 31 of the 34 schools had reached at least a moderate level of alignment in terms of implementation on both of the key program dimensions and that 20 of the programs were well aligned with the program models on both implementation dimensions. Classroom observations conducted during the study's second year used the same protocols and process as those conducted in the first year of implementation, except that only one observer visited the classrooms rather than two.

¹²Twenty-five of these teachers taught the ERO courses the entire year. Two of the returning teachers replaced other ERO teachers in the middle of the first year, and thus returned the second year having taught the ERO course less than a full year.

Chapter 4

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

In addition to examining the fidelity with which the sites participating in the Enhanced Reading Opportunities (ERO) study implemented the models of the two supplemental high school literacy programs — Reading Apprenticeship Academic Literacy and Xtreme Reading — the evaluation also includes an assessment of how much students participated in the ERO classes and whether they participated in other literacy support services either in or outside school. The evaluation team collected data about the frequency with which the ERO classes met and about whether and how often students attended. These data provide an indication of the overall “dosage” of the ERO interventions that students in the ERO group received during the first year of the study. The impact of the ERO programs will be a function, in part, of how much exposure the ERO students have to the classes throughout the school year. These data also provide an indication of whether students in the non-ERO group inadvertently enrolled in the ERO classes and thus diluted the overall contrast in literacy services received by students in the ERO and non-ERO groups.

The ERO evaluation team also collected data on the frequency with which students participated in classes or tutoring services that aimed to improve students’ reading and writing skills. Specifically, the student follow-up survey asked several questions about the frequency and duration with which students participated in such activities either in school or outside school. These data are available for students in both the ERO and the non-ERO groups and are intended to capture participation in both the ERO classes and other literacy support programs and services. They provide a measure of the difference in exposure to supplemental literacy support services between the ERO and non-ERO groups — which is a key factor in whether the ERO programs offer a contrast to the services that would otherwise be available.

This chapter discusses the following key findings:

- The ERO classes began an average of six weeks after the start of the school year and operated for an average of just over seven and a half months of the nine-month school year.
- More than 95 percent of the students in the ERO group enrolled in the ERO classes, and 91 percent were still attending the classes at the end of the school year.

- Students attended 83 percent of the scheduled ERO classes each month, and they received an average of just over 11 hours of ERO instruction per month.
- There were no systematic differences in ERO class enrollment and attendance rates between schools using Reading Apprenticeship and those using Xtreme Reading.
- Students who were randomly assigned to the study's ERO group reported a much higher frequency of participation in supplemental literacy services (in ERO classes and otherwise) than students in the non-ERO group. Although the largest difference occurred in a school-based literacy class, ERO students were also more likely to report working with a tutor in and outside school and attending a literacy class outside school.

In general, the ERO classes served as the primary source of literacy support services for students in the study sample. For students in the study's ERO group, the ERO classes substituted for a scheduled elective class — such as a career/technical education class, an arts class, a physical education or health class, or a foreign language class — and not one of the core-content classes: English/language arts, history/social studies, mathematics, and science. The ERO classes were not a source of literacy support for non-ERO students. Seven out of the 1,428 students in the non-ERO group enrolled in the ERO classes. Also, given that the ERO teacher at each school taught no other classes other than the ERO class, the only way for non-ERO students to receive ERO instruction was through enrollment in the ERO classes.

Student Enrollment and Attendance in the ERO Classes

The amount of ERO instruction that students receive is a function of program duration and student attendance. The longer the duration of the program, the greater the opportunity students have to participate in the ERO classes. The more often students attend, the more ERO instruction they will be exposed to. Following is an overview of findings from the evaluation's analysis of program duration and attendance.

Program Duration

The ERO programs were designed to operate for the full school year and to provide students with approximately nine months of supplemental literacy instruction. In fact, the ERO classes began an average of six weeks after the start of the 2005-2006 school year, ranging from

three to ten weeks across the 34 high schools.¹ The delayed start-up of the classes meant that the ERO programs operated for an average of just over seven and a half months rather than the full nine months of the school year.² This ranged from six and a half months in one school to eight and a half months in three schools. On average, across the participating high schools, students in Cohort 1 of the study sample had the potential to experience about 85 percent of the full planned Reading Apprenticeship and Xtreme Reading programs. Overall, during the first year of the project, 22 of the 34 participating high schools operated their ERO programs for more than seven and a half months.

Conducting the student recruitment and random assignment process at the start of the school year also meant that student class schedules had to be changed for the individuals assigned to the ERO group. This disrupted ERO students as they were pulled from elective classes and placed into ERO classes. In interviews with the study team, many of the ERO teachers reported that it took several days for students to settle in to their new schedules and adjust to the new expectations and routines.

Student Enrollment and Attendance

As part of their responsibilities to the project, the ERO teachers were required to maintain and report to the study team daily attendance records for all students randomly assigned to the ERO group. They were also asked to determine whether chronically absent students were still enrolled in the ERO programs or had transferred to another school in the district. These data, along with information about the length of ERO class periods, provided the basis for calculating several measures of ERO enrollment and attendance. These measures are displayed in Table 4.1.³

Overall, nearly 96 percent of students in the ERO group attended at least one ERO class during the year, and approximately 91 percent were still attending ERO classes at the end of the school year. On average, students remained enrolled in the ERO programs for just over seven months during the school year. Table 4.1 shows that similar percentages of students enrolled in and remained in the Reading Apprenticeship and the Xtreme Reading classes.

¹Because the selection of districts to receive the special SLC grants did not occur until June 2005, the student recruitment process was delayed until the start of the 2005-2006 school year. This required between three and 10 weeks to complete.

²Each of the participating high schools was in session for approximately nine months, excluding vacations.

³The findings presented in Table 4.1 are based on attendance data for ERO group students in the follow-up respondent sample — the same sample as is used in the impact analysis for this report. The ERO enrollment and attendance findings for these students provide an assessment of the dosage of ERO program services that is associated with the impact findings discussed in Chapter 5. Note that all measures in Table 4.1 include students from the ERO group who never enrolled in the ERO classes and students who left the program during the school year. Zero values were included for these students during the periods when they were not enrolled in the programs.

The Enhanced Reading Opportunities Study

Table 4.1

**Attendance in ERO Classes,
Follow-Up Respondent Sample in the ERO Group**

Characteristic	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Ever attended an ERO class during the year (%)	95.5	94.9	96.0
Attending ERO classes at the end of the year (%)	91.2	91.0	91.4
Average daily attendance rate in ERO classes per month ^a (%)	82.7	81.7	83.6
Number of months ERO program was in operation	7.7	7.8	7.7
Average number of months attending ERO classes	7.1	7.1	7.1
Average number of hours ERO class met per month	13.6	13.5	13.7
Average number of hours student attended ERO class per month	11.3	11.2	11.5
Sample size	1,408	686	722

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study monthly attendance data.

NOTES: Tests of statistical significance were not performed.

^aThere were 64 students who never attended an ERO class, 35 students from Reading Apprenticeship schools and 29 students from Xtreme Reading schools. Excluding these students, the average daily attendance rate for the remaining students who attended at least 1 ERO class is 86.6 percent for all schools, 86.1 percent for Reading Apprenticeship schools, and 87.1 percent for Xtreme Reading schools.

The ERO programs were designed for an average of 3 hours and 45 minutes of class time per week (which is scheduled either as 45-minute classes each day or as 80- to 90-minute classes every other day). With an average of 20 days of school per month, the ERO classes were designed to provide students with approximately 15 hours of supplemental literacy instruction per month. Based on the attendance data provided by the ERO teachers, Table 4.1 shows that the ERO classes met for an average of 13.6 hours per month (approximately 3 hours and 25 minutes per week). On average, students in the ERO group attended 82.7 percent of the scheduled ERO classes each month. This amounts to an average of 11.3 hours of ERO instruction per month, or just under 3 hours per week.

Student Participation in Literacy Support Activities

A requirement of the ERO funding grants from ED was that the participating schools would not operate other supplemental literacy programs during the evaluation period. This was to ensure that the effectiveness of the ERO programs could be evaluated in a context where they were not being compared with similar interventions. School district officials were asked in their grant applications to affirm that none of the schools included in their grant applications were currently using or planning to implement supplemental adolescent literacy programs for their ninth-grade students.⁴ At the same time, students in both the ERO and the non-ERO group were free to seek out other literacy-related services on their own. In some cases, they found other adults in the school to provide tutoring; in other cases, students and their families sought out other classes or tutors outside school.

This section of the chapter examines the extent to which the availability of the ERO programs created a sharp contrast in ERO students' exposure to supplemental literacy services in and outside school compared to students in the study's non-ERO group. To the degree that students in the non-ERO group participated in supplemental literacy support services either in or outside school, the overall contrast with the ERO group's participation in the ERO classes would be reduced. Before turning to this analysis, the chapter first reviews the manner in which the ERO classes were inserted into students' course schedules and discusses the degree to which literacy instruction was embedded in the typical English/language arts classes in the participating high schools.

Elective Courses

The ERO class was intended to substitute for an elective class, rather than for a core academic class, in students' ninth-grade course schedules. Each of the participating high schools used scheduling models that allowed students to take seven or eight courses during the year. Four of these courses were academic requirements such as English/language arts (ELA), mathematics, history, and science, leaving three or four slots for elective classes. Even in high schools where one of those slots was filled with another required course like physical education or health, there were still two or three slots open for electives. Thus, the primary difference between the ERO group and the non-ERO group is that the ERO students had one of their elective classes replaced by the ERO class and that the non-ERO students remained in their elective classes. This section of the chapter discusses the nature of the classes taken by the non-ERO group. These constitute a primary feature of the "counterfactual" to the ERO classes.

⁴U.S. Department of Education (2005).

A review of class schedules for students in the ERO group confirmed that the ERO classes did not replace English/language arts or other required academic classes (mathematics, history, or science). All students in the ERO and non-ERO groups were enrolled in the academic classes that were required by the school or district. The class schedules for the non-ERO group show that alternatives to the ERO classes consisted of a wide array of ninth-grade elective classes, over three-quarters of which fell into four main categories. Twenty-five percent of these classes were in the subject area of career and technical education; 21 percent were in the visual and performing arts; 16 percent were in physical education/health; 15 percent were in a foreign language.⁵ That is, with few exceptions, students in the non-ERO group were *not* enrolled in the ERO classes and, instead, were enrolled in a variety of electives.⁶ However, because ERO students had room for one or more electives beyond the ERO class in their schedules, this same variety of elective classes included students from both the ERO and the non-ERO group. In these shared courses, though, students in the ERO group were underrepresented relative to students in the non-ERO group. In short, non-ERO students were typically enrolled in four or five core required classes and two or three elective classes, while ERO students were typically enrolled in the same four or five core required classes, one or two of the same elective classes, and the ERO class. The ERO class never substituted for one specific elective for all ERO students at any given school.

To demonstrate how ERO fits into student schedules, two examples are presented in Table 4.2. Between them, these examples represent the three most common types of variation in student schedules: the schedule model, the number of course slots within the schedule model, and the number of required courses. First, the two most commonly used schedule models in the 34 high schools were the *traditional bell schedule*, in which each class typically meets daily for 40 to 50 minutes (Example 1); and the *alternating (or A/B) block schedule*, in which each class meets for about 90 minutes every other day (Example 2). Second, since the modal number of course slots in the schools' schedule models was 8.0 and the mean was 7.7, one example reflects a schedule with seven course slots, and the other has eight course slots. Lastly, as noted above, some schools may have included another required course (for example, physical education or health) beyond the four core academic courses.

Interviews with elective teachers supplemented the data about elective courses that were obtained from student schedules. Specifically, these interviews provided data about whether the elective courses focused explicitly on teaching reading and writing skills, thus offering students a similar opportunity as the ERO classes. For the few non-ERO elective classes

⁵These figures are based on a more detailed analysis of the 319 elective courses listed on student schedules from 10 of the 34 ERO high schools, one school from each district.

⁶Seven out of 1,428 non-ERO students from the study sample were found to have enrolled in an ERO class.

The Enhanced Reading Opportunities Study

Table 4.2

Comparison of ERO and Non-ERO Student Schedules

Example 1: Traditional Bell Schedule, Seven Periods, Four Required Courses

Period	ERO Students	Non-ERO Students
1	English/Language Arts	English/Language Arts
2	Math	Math
3	Science	Science
4	Social Studies/History	Social Studies/History
5	<i>ERO</i>	<i>Elective</i>
6	Elective	Elective
7	Elective	Elective

Example 2: Alternating (A/B) Block Schedule, Eight Periods, Five Required Courses

Period	ERO Students		Non-ERO Students	
	Day A	Day B	Day A	Day B
1	English/Language Arts	Science	English/Language Arts	Science
2	Math	Social Studies/History	Math	Social Studies/History
3	Required course	<i>ERO</i>	Required Course	<i>Elective</i>
4	Elective	Elective	Elective	Elective

NOTE: These are not actual schedules but represent two types of schedules in ERO high schools. They are used to demonstrate how ERO fits into student schedules.

where reading and writing were taught more explicitly, there were both ERO and non-ERO students indicating that exposure to these types of literacy supports was distributed across both groups. That is, the enrollments in these courses did not exclusively represent one group or the other, nor were all students from one group or the other enrolled in these courses. Specifically, in three different high schools, three courses were identified that had explicit literacy instruction, and they enrolled an average of 10 non-ERO students and five ERO students. These are the only three of hundreds of elective courses taken by students in the non-ERO and ERO groups across the 34 high schools in the study that were judged to include explicit literacy instruction. Even here, the classes enrolled a small proportion of the non-ERO group, and they included similar numbers of ERO and non-ERO students.

English/Language Arts Instruction

ELA classes offered another venue where literacy instruction might occur beyond elective courses and different kinds of supplemental literacy services. Both ERO and non-ERO students were enrolled in ELA classes together, and they received the same amount of ELA instruc-

tion. Interviews were conducted with ELA teachers that investigated the nature of the ELA instruction at the 34 high schools, with particular focus on assessing whether literacy-rich ELA instruction was already occurring. Because the ELA instruction was the same for ERO and non-ERO students, literacy-rich ELA instruction would not cause differences between those two groups of students but would, rather, possibly decrease the potential value added by the ERO classes. In interviews with members of the study team, ELA teachers across all of the participating schools indicated that their classes were comprised primarily of exposing students to different literary genres and some instruction in grammar and composition. While there was regular use of reading and writing activities, the instruction was literature-based and was not focused explicitly on improving reading and writing skills with the intensity or specificity of the ERO classes.

Overall, the support for building students' literacy skills available in the ninth-grade year to students in the non-ERO group through ELA and elective classes was not comparable in focus and intensity to that provided by the ERO classes. The ERO classes offered a strong contrast to the experiences of the non-ERO students, and they were different from other elective classes in their focus on literacy instruction. While the ERO programs were not taught in a literacy vacuum (that is, all students had reading and writing activities as part of their courses), they did provide support to students that was different and more intensive than what they typically received.

Student Participation in Supplemental Literacy Support Activities

The student follow-up survey included items aimed at determining the amount of extra literacy support that students received during the school year, beyond their regular English/language arts class. The survey asked about four categories of extra literacy help: classes in school, classes outside school, an adult tutor in school, and an adult tutor outside school. The first category describes such supports as the ERO courses. This item essentially provides an opportunity for ERO students to report on their attendance in the ERO classes, and for non-ERO students to report on their participation in literacy support activities that would be most similar to or “competitive” with ERO. The other three categories of activities cover other ways in which students might receive help with their reading and writing skills.

The survey questions asked all students about how long (duration) and how often (frequency) they participated in each of the four categories of activities. For example, a student who attended a “help” session every day for the full school year was projected to have attended approximately 180 sessions (about 20 days per month for nine months, or the typical number of days in a school year). Similarly, a student who reported attending twice per week for a semester was projected to have attended about 36 sessions (eight days per month for about four and a half months).

Table 4.3 provides the average levels of student participation in these four types of supplemental literacy support activities. The table also includes estimates of the differences in participation between the ERO and non-ERO groups.

The comparisons of the two groups provide an indication of the increase in literacy instruction and support that the ERO programs produce over and above what students would be exposed to without the programs. Reflecting their participation in the ERO program, students in the ERO group participated in a school-based literacy class six times more than students in the non-ERO group. It should be noted, however, that students in the non-ERO group did report receiving some exposure to a literacy class in school, though only a handful of non-ERO students across all the high schools ever enrolled in an ERO class. Table 4.3 also shows that students in the ERO group also reported higher levels of participation in tutoring sessions and in literacy classes outside school than students in the non-ERO group.

The Enhanced Reading Opportunities Study

Table 4.3

**Participation in Supplemental Literacy Support Activities,
Cohort 1 Follow-Up Respondent Sample**

Outcome	ERO Group	Non-ERO Group	Impact	Impact Effect Size	P-Value for the Difference
<u>All schools (number of sessions)</u>					
School-based literacy class	63.2	11.4	51.8 *	1.44 *	0.000
School-based adult tutor	22.0	8.2	13.9 *	0.46 *	0.000
Outside-school literacy class	5.5	2.5	3.0 *	0.20 *	0.001
Outside-school adult tutor	8.6	5.5	3.1 *	0.13 *	0.011
Sample size	1,410	1,002			
<u>Reading Apprenticeship schools (number of sessions)</u>					
School-based literacy class	64.1	11.6	52.5 *	1.46 *	0.000
School-based adult tutor	21.0	8.5	12.6 *	0.42 *	0.000
Outside-school literacy class	5.1	3.6	1.5	0.10	0.302
Outside-school adult tutor	8.8	7.1	1.7	0.07	0.356
Sample size	689	455			
<u>Xtreme Reading schools (number of sessions)</u>					
School-based literacy class	62.3	11.0	51.3 *	1.43 *	0.000
School-based adult tutor	23.0	8.0	14.9 *	0.49 *	0.000
Outside-school literacy class	5.8	1.4	4.4 *	0.29 *	0.000
Outside-school adult tutor	8.5	4.1	4.3 *	0.19 *	0.007
Sample size	721	547			

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-based class standard deviation = 35.924; school-based tutor standard deviation = 30.240; outside-school class standard deviation = 14.896; outside-school tutor standard deviation = 23.027).

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

For each of the above measures, data are missing for no more than 6 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

Chapter 5

Early Impacts on Student Reading Achievement and Reading Behaviors

The primary focus of the Enhanced Reading Opportunities (ERO) evaluation is to assess the impact of supplemental literacy interventions on adolescent students' reading comprehension skills and behaviors and on their overall academic performance during high school. The early impact analysis presented in this report addresses two questions that pertain to the first year in which the ERO programs were being implemented and to their effects for ninth-grade students at the end of the year in which they were enrolled in the programs:¹

- What is the impact of supplemental literacy programs on ninth-grade students' reading comprehension as measured by standardized test scores for reading comprehension and reading vocabulary?
- What is the impact of supplemental literacy programs on ninth-grade students' vocabulary and on their reading behaviors as measured by self-reported information about how much students read and whether they use specific reflective reading strategies?

Because the study's two supplemental literacy programs — Reading Apprenticeship Academic Literacy and Xtreme Reading — focus on producing immediate improvements in students' reading comprehension ability, the early impact analysis presented in this report places a higher priority on the first question above. Each of the programs also endeavors to enhance students' vocabulary and their interest in reading both in and outside school and to increase their use of strategies that are characteristic of proficient readers. For this reason, the analysis also examines impacts on vocabulary test scores and on three measures of students' reading behaviors. As discussed in Chapter 2, measures of students' reading comprehension and vocabulary skills are drawn from their performance on the Group Reading Assessment and Diagnostic Examination (GRADE) administered at the end of their ninth-grade year. The measures of reading behavior were developed from the follow-up survey that was administered to students in the study sample at the end of their ninth-grade year.

This chapter first presents early impact findings for all 34 of the high schools in the evaluation. The results that are pooled across the two programs selected for the demonstration

¹Subsequent reports will also examine impacts on a range of longer-term outcomes, including performance on standardized state tests, credits earned toward graduation, daily attendance, grade-to-grade promotion rates, and dropout rates.

provide evidence about the effectiveness of the two supplemental literacy interventions selected by the expert panel for this project as a class of interventions. The chapter then presents findings for each of the two ERO programs separately. Although Reading Apprenticeship and Xtreme Reading share overarching goals for adolescent literacy development and share many instructional principles, these results provide evidence about whether their differences in operating strategies resulted in different patterns of impacts.

The chapter also summarizes findings for subgroups of students defined by pre-random assignment background characteristics, including their baseline reading test scores, whether they had repeated an earlier grade, and whether a language other than English is spoken at home.

The chapter ends with an exploration of variation in impacts across two subgroups of schools in the study. The implementation of the ERO programs in one group of schools was classified as at least moderately aligned with the program models — as defined in Chapter 3 — and the schools were able to operate their ERO programs for more than seven and a half months (the average for the sample as a whole). The implementation of the ERO programs in the other group of schools were classified as poorly aligned with their program models or they operated for seven and a half months or less. It is not possible to conclude definitively that differences in impacts between these two groups of schools were *caused* by differences in their early start-up experiences. Rather, this analysis represents an exploration of impacts under conditions that were more like those intended by the program developers.

The chapter discusses the following key findings:

- Overall, the ERO programs produced a positive and statistically significant impact on reading comprehension test scores, with an effect size of 0.09 standard deviation. This impact corresponds to an improvement from the 23rd percentile nationally, as represented by the average scores for students in the non-ERO group, to the 25th percentile nationally, as represented by the average scores for students in the ERO group.
- Despite the positive impact on reading comprehension test scores, almost 90 percent of students in the study sample who enrolled in the ERO programs were still reading below grade level at the end of the ninth grade.
- Although they are not statistically significant, the magnitudes of the impact estimates on reading comprehension test scores for each literacy intervention are the same as those for the full study sample.
- The ERO programs did not produce statistically significant impacts on vocabulary test scores.

- The ERO programs exhibited a mix of positive and negative impacts on the measures of reading behavior, but these are not statistically significant.
- Positive impacts on reading comprehension were concentrated among schools whose implementation of the ERO programs was at least moderately aligned with the program models and schools that were able to operate their ERO programs for more than seven and a half months.

Early Impacts on Reading Achievement

The ERO study assesses the impact of supplemental literacy interventions of the type represented by Reading Apprenticeship and Xtreme Reading. As such, the analysis focuses first on impacts that are pooled across both interventions and all sites in the study sample. Thus, in pooling the sample across all schools in the study, the analysis has sufficient power to detect statistically significant impacts that are somewhat smaller than those that can be detected for each ERO program separately. At the same time, the study was designed to ensure adequate statistical power for policy-relevant impact estimates from each intervention separately. The primary measure of reading achievement for this study is students' scores on the GRADE reading comprehension assessment. A secondary measure of students' reading achievement is their scores on the GRADE vocabulary assessment.

- **Overall, the ERO programs produced a positive and statistically significant impact on reading comprehension (0.90 standard score point, which corresponds to an effect size of 0.09 standard deviation).**

The first row in Table 5.1 shows that, averaged across all 34 participating high schools, the ERO programs improved reading comprehension test scores by 0.9 standard score point and that this impact is statistically significant (p-value is less than or equal to 5 percent). Expressed as a proportion of the overall variability of test scores for students in the non-ERO group, this represents an effect size of 0.09 (or 9 percent of the standard deviation of the non-ERO group's test scores). Table 5.1 also shows that this impact corresponds to an improvement from the 23rd percentile nationally, as represented by the average scores for students in the non-ERO group, to the 25th percentile nationally, as represented by the average scores for students in the ERO group.

Figure 5.1 places this impact estimate in the context of the actual and expected change in the ERO students' reading comprehension test scores from the beginning of ninth grade to the end of ninth grade. The bottom section of the bar shows the average reading comprehension test score for students in the ERO group at the beginning of their ninth-grade year. This average of 85.9 standard score points corresponds, approximately, to a grade equivalent of 5.1 and indicates an average reading level for students nationally at the start of fifth grade. This marks the

The Enhanced Reading Opportunities Study

Table 5.1

**Impacts on Reading Achievement,
Cohort 1 Follow-Up Respondent Sample**

Outcome	ERO Group	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	89.2	0.9 *	0.09 *	0.019
<i>Corresponding grade equivalent</i>	6.1	5.9			
<i>Corresponding percentile</i>	25	23			
Reading vocabulary					
Average standard score	93.4	93.2	0.3	0.03	0.472
<i>Corresponding grade equivalent</i>	7.7	7.7			
<i>Corresponding percentile</i>	32	31			
Sample size	1,408	1,005			

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 2006 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

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.

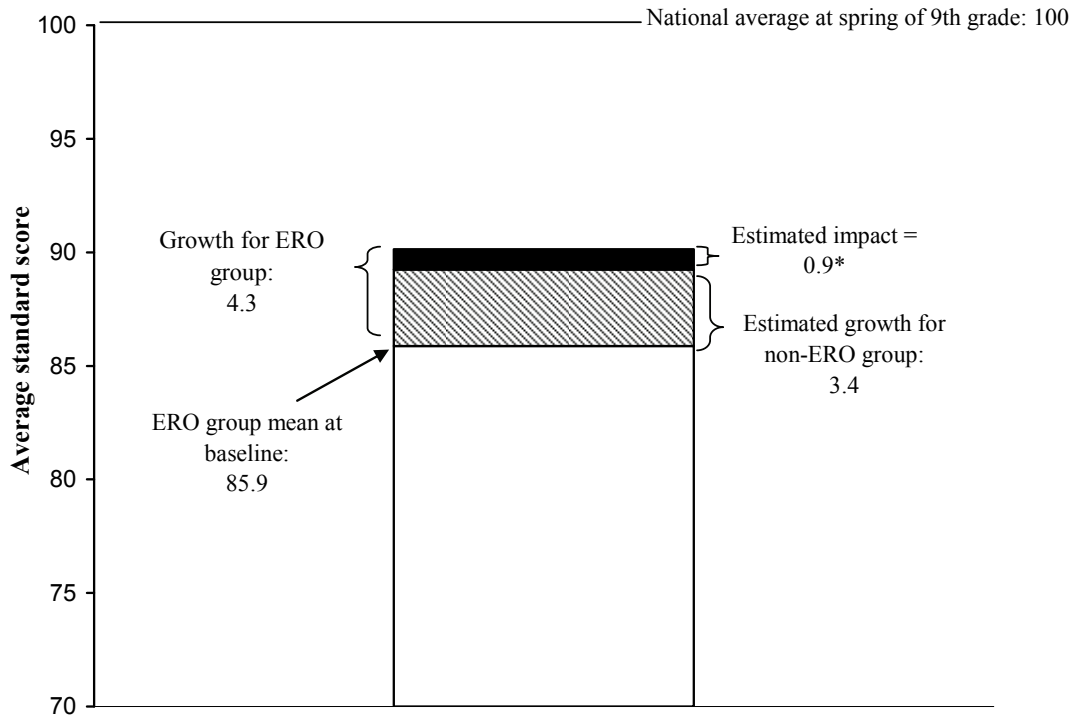
The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (reading comprehension = 10.458; reading vocabulary = 10.505).

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.

The Enhanced Reading Opportunities Study

Figure 5.1
Impacts of Reading Comprehension,
Cohort 1 Follow-Up Respondent Sample



SOURCES: MDRC calculations from the Enhanced Reading Opportunities Study baseline and follow-up GRADE assessments; American Guidance Service, *Group Reading Assessment and Diagnostic Evaluation: Teacher's Scoring and Interpretive Manual, Level H*.

NOTES: The baseline GRADE assessment was administered in the fall of 2005 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 2006 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.

starting point for measuring both the observed growth in their reading achievement through the end of their ninth-grade year and their expected growth to be estimated through the test scores of the non-ERO group at the end of ninth grade.

Together, the bottom two sections of the bar in Figure 5.1 show the estimated reading comprehension test scores of students in the non-ERO group at the end of their ninth-grade year. The middle section of the bar, therefore, represents the growth in test scores experienced by the non-ERO group. This growth of 3.4 points 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 ERO impact on reading comprehension test scores. Thus, the impact of the ERO programs represents a 26 percent improvement over and above what the ERO group would have achieved if they had not had the opportunity to attend the ERO classes.² From this perspective, the ERO programs produced more progress on reading comprehension than the gains expected for this sample of students had they not been selected for the programs.

Together, the top two sections of the bar in Figure 5.1 indicate that students in the ERO group improved by an average of 4.3 standard score points over the course of their ninth-grade year. Thus, the impact of the ERO programs accounts for 21 percent of the average test score improvement experienced by the ERO group.³

The solid line at the top of Figure 5.1 shows the national average (100 standard scale 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. Despite the program impact, therefore, students' reading comprehension scores still lagged nearly 10 points below the national average for performance on GRADE reading comprehension for students at the end of their ninth-grade year. In fact, almost 90 percent of the students in the ERO group had reading comprehension scores that were below grade level, and 76 percent had scores that were two or more years below grade level.

- **Although the difference is not statistically significant, vocabulary test scores for students in the ERO group were estimated to be 0.3 standard score point higher than those for the non-ERO group.⁴**
- **Estimated impacts on reading comprehension and vocabulary test scores for each ERO program are not statistically significant.**

²This was calculated by dividing the impact (0.9 standard score point) by the average improvement of the non-ERO group (3.4 standard score points).

³This was calculated by dividing the impact (0.9 standard score point) by the average improvement of the ERO group (4.3 standard score points).

⁴The ERO study did not include a vocabulary test at baseline. As a result, it is not possible to place the impacts on vocabulary in the context of changes that occurred over the course of students' ninth-grade year.

Table 5.2 shows that the impacts on reading comprehension for both Reading Apprenticeship and Xtreme Reading are of the same magnitude as that found for the full sample of schools in the study. However, neither of these results is statistically significant. The table also shows that neither ERO program produced a statistically significant impact on vocabulary test scores.

Early Impacts on Students' Reading Behaviors

As noted in Chapter 2, the Enhanced Reading Opportunities Student Follow-Up Survey was administered at the same time as the follow-up GRADE assessment, at the end of the students' ninth-grade year. The impact analysis presented in this chapter focuses on three measures of students' reading behavior that were derived from the survey: amount of school-related reading, amount of non-school-related reading, and use of reflective reading strategies.⁵ Table 5.3 presents early findings on the ERO programs' average impact on these three measures. Table 5.4 presents these results separately for each of the two ERO programs.

- **Overall, the ERO program impacts on the reading behavior measures were not statistically significant.**

Each of the two supplemental literacy programs seeks to motivate students to read more. They do this both by providing opportunities to read and discuss what they read in the ERO classes and by providing classroom libraries and assigning texts for students to read at home. The goal is to expose students to a wide range of reading opportunities, while building the strategies that proficient readers use and thereby stimulating students' interest in reading more both for school and for their own enjoyment.

Table 5.3 shows that, across all 34 high schools, the amount of reading that students in the ERO group reported is greater than that of students in the non-ERO group. Neither of these results is statistically significant. The impact on students' reports of using reflective reading strategies is nearly zero.

Table 5.4 shows the impacts on reading behaviors separately for each ERO program. Although the bottom panel of Table 5.4 indicates that Xtreme Reading produced a positive and statistically significant impact on the amount of school-related reading that students reported, this result should be interpreted cautiously. As noted in Chapter 2, the analyses include qualifying statistical tests aimed at assessing the robustness of multiple impacts within the reading behavior measurement domain. The qualifying tests examine the estimated impact on a composite index of reading behaviors for each ERO program separately and a test of whether the difference in

⁵A list of the survey items used to create these three measures is presented in Appendix A.

The Enhanced Reading Opportunities Study

Table 5.2

Impacts on Reading Achievement,
Cohort 1 Follow-Up Respondent Sample,
by Program

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Reading Apprenticeship schools</u>					
Reading comprehension					
Average standard score	89.8	88.9	0.9	0.09	0.097
<i>Corresponding grade equivalent</i>	6.1	5.9			
<i>Corresponding percentile</i>	24	23			
Reading vocabulary					
Average standard score	93.2	92.8	0.5	0.05	0.393
<i>Corresponding grade equivalent</i>	7.7	7.7			
<i>Corresponding percentile</i>	31	31			
Sample size	686	454			
<u>Xtreme Reading schools</u>					
Reading comprehension					
Average standard score	90.5	89.6	0.9	0.09	0.090
<i>Corresponding grade equivalent</i>	6.2	6.0			
<i>Corresponding percentile</i>	25	24			
Reading vocabulary					
Average standard score	93.6	93.5	0.1	0.01	0.846
<i>Corresponding grade equivalent</i>	7.8	7.8			
<i>Corresponding percentile</i>	32	32			
Sample size	722	551			
Difference in Impacts Between Programs					
			Difference in Impacts	Difference in Impact Sizes	P-Value for Difference
<u>Reading Apprenticeship minus Xtreme Reading</u>					
Reading comprehension standard score			0.0	0.00	0.962
Reading vocabulary standard score			0.4	0.04	0.664

(continued)

Table 5.2 (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 2006 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

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 = 10.458; reading vocabulary = 10.505).

A two-tailed t-test was applied to the impact estimate and to the difference in impacts. 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.

impacts between the two groups of schools is statistically significant.⁶ These tests indicate that neither ERO program produced a statistically significant impact on the composite index that was created to capture the three reading behavior measures. Also, the difference in the impacts on the composite index between the two programs was not statistically significant. As a result, the one statistically significant result presented in Table 5.4 should be interpreted cautiously.

Early Impacts for Subgroups of Students

While all students in the study sample had baseline reading comprehension skills between the fourth- through seventh-grade level at the start of ninth grade, the ERO study sample includes a diverse population of students. With this diversity in mind, the ERO evaluation was designed to allow for the estimation of impacts for key subgroups of students who face especially challenging barriers to literacy development and overall performance in high school. For example, prior research has shown that especially low literacy levels, evidence of failure in prior grades, and having English as a second language are powerful predictors of school success.⁷

This section of the chapter and Appendix H examine variation in ERO program impacts for subgroups of students defined by their baseline reading comprehension test scores, whether

⁶See Appendix E, Appendix Table E.3, for the results of these qualifying tests.

⁷Roderick (1993); Fine (1988).

The Enhanced Reading Opportunities Study

Table 5.3

**Impacts on Reading Behaviors,
Cohort 1 Follow-Up Respondent Sample**

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>All schools</u>					
Amount of school-related reading (prior month occurrences)	44.2	43.4	0.8	0.02	0.669
Amount of non-school-related reading (prior month occurrences)	27.3	26.0	1.3	0.04	0.315
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	-0.01	0.849
Sample size	1,410	1,002			

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 43.867; non-school-related reading standard deviation = 31.834; use of reading strategies standard deviation = 0.670).

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.

For each of the above measures, data are missing for no more than 5 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

they were overage for the ninth grade, and whether a language other than English was spoken in their homes. As reported in Chapter 2 (see Table 2.4), 36 percent of the study sample had baseline test scores that indicated reading levels that were four to five years below grade level at the start of ninth grade, and another 28 percent were reading from three to four years below grade level. Also, over a quarter of the students in the study sample were overage for the ninth grade (that is, they were age 15 years or older at the start of ninth grade), which is used to indicate that

The Enhanced Reading Opportunities Study
Table 5.4
Impacts on Reading Behaviors,
Cohort 1 Follow-Up Respondent Sample,
by Program

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Reading Apprenticeship schools</u>					
Amount of school-related reading (prior month occurrences)	43.8	48.3	-4.5	-0.10	0.100
Amount of non-school-related reading (prior month occurrences)	26.8	27.6	-0.8	-0.02	0.672
Use of reflective reading strategies (4-point scale)	2.6	2.7	0.0	-0.03	0.600
Sample size	689	455			
<u>Xtreme Reading schools</u>					
Amount of school-related reading (prior month occurrences)	44.5	39.2	5.3 *	0.12 *	0.029
Amount of non-school-related reading (prior month occurrences)	27.7	24.6	3.1	0.10	0.081
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	0.02	0.779
Sample size	721	547			
<u>Difference in Impacts Between Programs</u>					
			Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>Reading Apprenticeship minus Xtreme Reading</u>					
Amount of school-related reading			-9.8 *	-0.22 *	0.007
Amount of non-school-related reading			-3.9	-0.12	0.133
Use of reflective reading strategies			0.0	-0.05	0.566

(continued)

Table 5.4 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 43.867; non-school-related reading standard deviation = 31.834; use of reading strategies standard deviation = 0.670).

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

For each of the above measures, data are missing for no more than 5 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

a student was retained in a prior grade.⁸ Approximately 45 percent of the students in the sample lived in households where a language other than English was spoken.

Table 5.5 provides a summary of impact findings for the subgroups of students defined by their baseline reading comprehension test scores, whether they were overage for the ninth grade, and whether a language other than English was spoken in their homes.⁹ In general, the table indicates that the ERO programs produced positive and statistically significant impacts on reading comprehension test scores for two of the subgroups and on vocabulary test scores for one of the subgroups. Nevertheless, the composite qualifying statistical test for the multiple hypothesis tests reflected in the table indicates that the overall variation in impacts across the subgroups is not statistically significant (F-statistic = 0.865; p-value = 0.534). Also, the difference in impacts between subgroups was not statistically significant.¹⁰

The first column in Table 5.5 shows the impact on reading comprehension test scores in effect size units. It indicates that the ERO programs produced positive and statistically significant impacts on reading comprehension test scores for students who were overage for grade and for students from multilingual families. However, the difference between these impacts and those for their counterpart subgroups of students are not statistically significant. As a result, although the

⁸National Center for Education Statistics (1990).

⁹Appendix Tables H.1 through H.6 in Appendix H provide the outcome levels for the ERO and non-ERO groups, the estimated impacts, impact effect sizes, and p-values for the estimates presented in Table 5.5. The tables in Appendix H also show the difference in estimated impacts across subgroups and p-values of these differences.

¹⁰See Appendix H. Also, as noted in Chapter 2, 423 students with baseline reading test scores that were not within the target range intended for the study were not included in the impact analysis for this report. Sensitivity tests of the impact estimates indicate that the findings are not sensitive to the inclusion of these students.

The Enhanced Reading Opportunities Study
Table 5.5

Impact Effect Sizes for Student Subgroups

	Number of Students	Reading Comprehension Impact		Vocabulary Impact		Amount of School-Related Reading Impact		Amount of Non-School-Related Reading Impact		Use of Reflective Reading Strategies Impact	
		Effect Size	P-Value	Effect Size	P-Value	Effect Size	P-Value	Effect Size	P-Value	Effect Size	P-Value
<u>Baseline reading comprehension</u>											
6.0-7.0 grade equivalent	855	0.10	0.107	0.12 *	0.040	0.02	0.760	0.11	0.126	-0.06	0.376
5.0-5.9 grade equivalent	680	0.08	0.274	-0.06	0.401	0.06	0.430	0.05	0.526	0.05	0.471
4.0-4.9 grade equivalent	878	0.08	0.233	-0.02	0.729	0.00	0.998	-0.03	0.691	0.00	0.956
<u>Overlap for grade^a</u>											
Students overage for grade	644	0.19 *	0.007	0.09	0.221	0.04	0.667	0.10	0.253	-0.03	0.676
Students not overage for grade	1,769	0.05	0.267	0.00	0.992	0.02	0.718	0.02	0.647	-0.01	0.876
<u>Language spoken at home</u>											
Students from multilingual families	1,133	0.12 *	0.027	0.10	0.072	0.12	0.052	0.12 *	0.031	-0.03	0.664
Students from English-only families	1,280	0.07	0.181	-0.03	0.512	-0.09	0.140	-0.05	0.387	-0.01	0.908

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

NOTES: Appendix H provides detailed information about each of the student subgroup impact estimates, including outcome levels for the ERO and non-ERO groups, impact estimates, p-values, and differences in impacts among subgroups.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (reading comprehension = 10.458; reading vocabulary = 10.505; school-related reading = 43.867; non-school related reading = 31.834; use of reading strategies = 0.670).

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.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

ERO programs produced a statistically significant impact on reading comprehension test scores for these subgroups of students, the analysis does not provide adequate confidence to conclude that the programs “worked better” for those students than it did for other subgroups of students.

The second column of Table 5.5 shows that the ERO programs produced a positive and statistically significant impact for students with baseline reading comprehension test scores that fell between the 6.0 and 7.0 grade equivalent. The difference between this impact and those for the other two test score subgroups is not statistically significant.

The far-right columns of Table 5.5 summarize the impacts on the reading measures for each of the subgroups. They indicate that the ERO programs produced a positive and statistically significant impact on the amount of non-school-related reading reported by students from multilingual families. In addition, the difference in impacts on school-related reading between students in multilingual families and students from English-only families is statistically significant. The qualifying tests that were conducted to account for the multiple hypothesis tests, however, indicate that the ERO program impacts on the composite index that was created to capture the three reading behavior measures are not statistically significant. Thus, the single statistically significant impact on reading behaviors in Table 5.5 should be interpreted cautiously.

The Relationship Between Early Impacts and First-Year Implementation Issues

This section of the chapter explores the variation in impacts of the ERO programs under conditions that were more or less proximal to those intended at the outset of the study and, as noted in Chapter 3, are more prevalent in the study’s second year than they were in the study’s first year. Specifically, it examines impacts for subgroups of the participating high schools that were defined by the degree to which they were able to achieve two implementation milestones during the first year of the study: whether they reached at least a moderate level of implementation fidelity (as defined in Chapter 3) and whether they were able to operate for more than seven and a half months (the average for the sample). The 15 schools that were able to reach both of these thresholds were deemed to have had a first-year start-up experience that was more in line with the original intent of the project than schools that did not meet these thresholds.

It is important to note that the analyses presented in this section of the chapter are exploratory and are not able to establish causal links between these early implementation milestones and variation in estimated impacts on student’s reading achievement across the sites. A variety of other program and school characteristics — not examined in the analyses presented here — may also be associated with differences in impacts across the schools. As an exploratory analysis, it is also not appropriate to extrapolate from these findings to predict the impact of the ERO programs in the second year of the project.

The exploration of relationships between impacts and first-year implementation challenges proceeds in three stages. The first stage provides an assessment of overall variation in impacts on reading comprehension test scores across the 34 participating schools. To the degree that there is variation in impacts across the sites, the overall average may be masking important differences in the effectiveness (or lack of effectiveness) of the ERO programs under some conditions. The second stage explores two sets of relationships: (1) the relationship between impacts and the implementation fidelity ratings and (2) the relationship between impacts and program implementation duration. The third stage combines the two indicators of first-year implementation challenges and presents impacts for two groups of sites based on whether they encountered serious problems either with implementation fidelity or with program duration during the first year of the study.

Overall Variation in ERO Impacts across Schools

Figure 5.2 illustrates the variation in estimated program impacts on reading comprehension scores across the 34 participating high schools.¹¹ For each school and for the overall average, the figure displays mean impact estimates (represented by the squares) and the 95 percent confidence intervals around the mean impact estimates (represented by the lines extending above and below the squares.) Here, the wider the confidence interval, the broader the margin of error and the greater the uncertainty about the impact estimate. Confidence intervals that do not include zero are statistically significant (p-value is less than or equal to 5 percent). The school-by-school impact estimates range from an ERO program producing a reduction in reading comprehension test scores of 7.1 standard score points to an ERO program producing an increase of 5.9 standard score points. In all, 23 estimates are positive, and 11 are negative; 16 estimates are smaller than the full-sample average, and 18 estimates are about the same or larger. Only five of the school-level impact estimates are statistically significant.

The variation in estimated impacts displayed in Figure 5.2 overstates the variation in true impacts, however, because a large portion of the variation in estimated impacts is due to estimation error. In other words, many of the estimates in the figure appear to be highly negative or highly positive; yet, for all but five of the estimates, their confidence intervals include zero, which indicates that they cannot be distinguished reliably from zero. For example, the second-most-negative impact is -3.7 standard score points, but its confidence interval ranges from -7.8 to 0.4 standard score points.

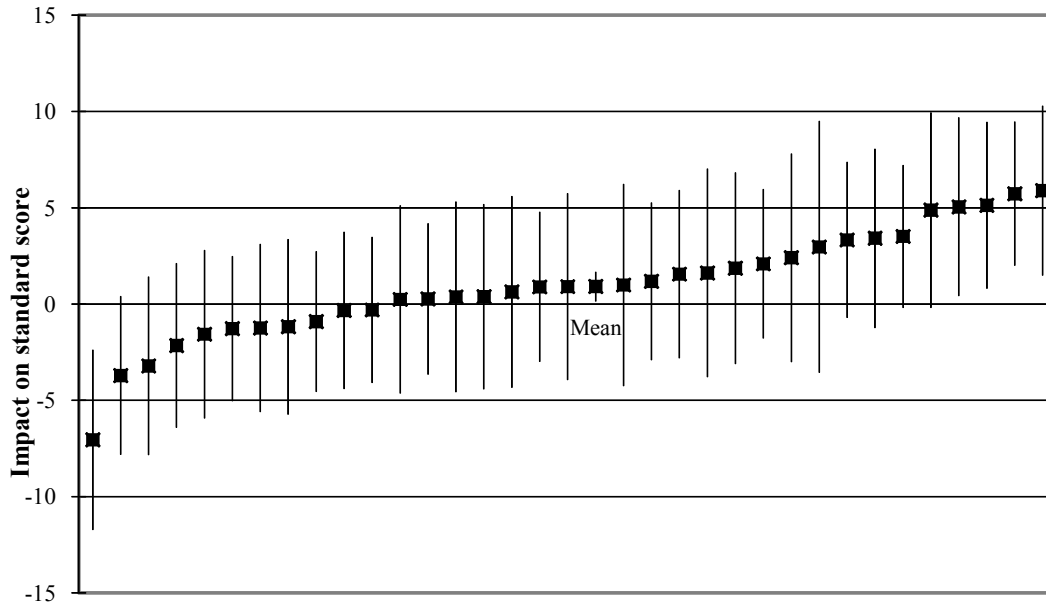
To examine variability in impacts across schools more systematically, a composite F-test was used to assess whether the school-level impacts on reading comprehension test scores

¹¹Estimated impacts are presented in numerical (ascending) order. See Appendix I for numeric values presented in Figure 5.2.

The Enhanced Reading Opportunities Study

Figure 5.2

Fixed-Effect Impact Estimates on Reading Comprehension, by School



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 2006 near the end of students' ninth-grade year.

The fixed-effects impact estimates are the regression adjusted impacts of the interaction between school and treatment 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.

are statistically equivalent. This test accounts for estimation error in school-level impacts and provides an indication of the confidence one might have that there is variation in true impacts across the schools. The results show that the p-value for the F-test is 0.013, indicating that the school-to-school variation in impacts is statistically significant and, thus, is unlikely to have occurred by chance.¹²

¹²See Appendix I for the results of this F-test.

Impacts Associated with Implementation Fidelity and Duration During the First Year

First, the analysis examines impacts for groups of schools defined by whether the implementation of their ERO programs was classified as well aligned, moderately aligned, or poorly aligned with their respective program models, as defined in Chapter 3. This analysis provides insight into the hypothesis that ERO programs could produce stronger impacts if they are able to create classroom learning environments and to develop instructional strategies that were deemed to be relatively closely aligned with the specifications of the program that they were using. The top panel of Table 5.6 provides a summary of impact findings for the subgroups of schools defined by the implementation-fidelity categories that are discussed in Chapter 3.¹³ The first column shows the estimated impact on reading comprehension test scores, and the second column shows the estimated impact on vocabulary test scores. The far-right pairs of columns show estimated impacts on the three reading behavior measures. All impacts estimates are presented in effect size units.

The top panel of Table 5.6 indicates that, on average, the 16 schools whose ERO programs had reached a well-aligned level of implementation fidelity on both the classroom learning environment and the comprehension instruction dimensions of their models produced positive, but not statistically significant, impacts on reading comprehension test scores. A similar impact is exhibited in the third row for the 10 schools whose ERO programs were found to have poorly aligned implementation fidelity on at least one of the two dimensions. Statistically significant impacts were found for the eight schools whose ERO programs reached at least a moderately aligned level of fidelity on both dimensions but were not able to reach an adequate level on at least one dimension. In fact, the difference in impacts on reading comprehension test scores between the schools in the moderately aligned fidelity category and schools in the poorly aligned fidelity category is statistically significant. This result should be interpreted cautiously, however, because a composite test indicates that overall variation in impacts across the three fidelity subgroups is not statistically significant.

The top panel of Table 5.6 also provides a test of the linear relationship between impacts and a continuous indicator of overall implementation fidelity.¹⁴ The result presented in

¹³Appendix Tables I.2 through I.7 in Appendix I provide the outcome levels for the ERO and non-ERO groups, the estimated impacts, impact effect sizes, and p-values for the estimates presented in Table 5.6. The tables in Appendix I also show the differences in estimated impacts across school subgroups and p-values of these differences.

¹⁴For the purposes of this analysis, an indicator was calculated as the average of the fidelity rating for the classroom learning environment dimension and the fidelity rating for the comprehension instruction dimension. A value was calculated for each school ranging from one to three and rounded to the nearest tenth. The interaction between this indicator and the treatment indicator was added to the impact estimation model. The parameter estimate for this interaction term indicates whether the ERO program impact increased or decreased as a linear function of the fidelity indicator.

The Enhanced Reading Opportunities Study
Table 5.6
Impact Effect Sizes, by First-Year Implementation Issues

	Number of Schools	Reading Comprehension		Vocabulary		Amount of School-Related Reading		Amount of Non-School-Related Reading		Use of Reflective Reading Strategies	
		Impact	P-Value	Impact	P-Value	Impact	P-Value	Impact	P-Value	Impact	P-Value
		Effect Size	Effect Size	Effect Size	Effect Size	Effect Size	Effect Size	Effect Size	Effect Size	Effect Size	Effect Size
<u>Fidelity of program implementation</u>											
Well-aligned implementation	16	0.06	0.260	-0.05	0.404	0.04	0.466	0.06	0.282	-0.02	0.778
Moderately aligned implementation	8	0.22 *	0.005	0.17 *	0.027	0.16	0.057	0.13	0.120	-0.07	0.362
Poorly aligned implementation	10	0.02	0.797	0.03	0.655	-0.13	0.115	-0.07	0.345	0.06	0.433
Continuous fidelity measure	34	-0.04	0.498	-0.12	0.075	0.02	0.794	0.03	0.637	-0.06	0.400
<u>Duration of program implementation</u>											
More than 8.0 months	7	0.16 *	0.039	-0.09	0.258	0.05	0.579	0.05	0.593	-0.06	0.482
7.6 to 8.0 months	15	0.10	0.081	0.06	0.239	-0.01	0.922	0.02	0.745	0.02	0.675
7.5 month or fewer	12	0.02	0.712	0.05	0.487	0.04	0.570	0.07	0.339	-0.02	0.754
Continuous duration measure	34	0.07	0.351	-0.09	0.246	-0.01	0.896	-0.04	0.682	0.00	0.968
<u>First-year implementation issues</u>											
Moderately or well-aligned and longer duration	15	0.17 *	0.002	0.01	0.848	0.11	0.065	0.10	0.075	0.01	0.887
Poorly aligned or shorter duration	19	0.01	0.811	0.04	0.412	-0.07	0.250	-0.02	0.744	-0.02	0.695

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

NOTES: Appendix I provides detailed information about each of the program implementation subgroup impact estimates, including outcome levels for the ERO and non-ERO groups, impact estimates, p-values, and differences in impacts among subgroups.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (reading comprehension = 10.458; reading vocabulary = 10.505; school-related reading = 43.867; non-school-related reading = 31.834; use of reading strategies = 0.670).

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.

Table 5.6 indicates that the linear relationship between impacts and this overall fidelity indicator is not statistically significant.

Finally, the top panel of Table 5.6 indicates that, with the exception of vocabulary, impacts on other outcomes across the groups of sites are not statistically significant. Although the ERO programs in the moderately aligned fidelity category of schools produced a positive and statistically significant impact on vocabulary test scores, the difference in impacts across the subgroups is not statistically significant.

The analysis now turns to an examination of impacts for subgroups of schools defined by how long they were able to implement their ERO programs during the first year of the study. The length of program operation encompasses two first-year implementation challenges. First, delays in the start-up of the ERO programs meant that students randomly assigned to the ERO programs had already spent between three and 10 weeks enrolled in a regular elective class that they would have to leave in order to enroll in an ERO class. Rescheduling them into the ERO class was disruptive and required that they acclimate themselves to a new teacher and set of classroom routines. Second, the variation in the start-up delays meant that different amounts of time were available for teachers to cover the course curricula for the ERO programs and for ERO students to receive exposure to the ERO activities and materials that were planned by the developers.

The middle panel of Table 5.6 shows estimated impacts for three groups of sites: those that were able to operate for more than eight months, those that were able to operate for more than seven and a half months but less than eight months, and those that were able to operate for seven and a half months or less. The designation of these groups of schools — particularly those at either end of the distribution — reflects key differences in the potential interaction between implementation and program start-up, or duration. Schools that experienced start-up delays of six weeks or more — and that could operate for only seven and a half months or less — may reflect the most disruptive start-up for students assigned to the ERO classes and had the shortest amount of time to cover the ERO curricula. On the other hand, while none of the programs was able to operate for the full school year, by operating for more than eight months of the nine-month school year, schools had the opportunity to expose their ERO students to nearly 90 percent of the ERO learning strategies and activities offered by their programs. Schools in the middle group were able to operate their ERO programs between seven and half and eight months.

The middle panel of Table 5.6 shows that the estimated impacts on reading comprehension are positive and statistically significant (effect size = 0.16 standard deviation and p-value = 0.039) for schools that operated for the longest period of time during the school year. Although the differences in impacts across the three subgroups of sites are not statistically significant, the table indicates that estimated impacts are smaller for schools with shorter operating periods (effect sizes = 0.10 and 0.02 standard deviation). Table 5.6 also provides a test of the linear rela-

tionship between impacts and a continuous indicator of the number of months of ERO program implementation.¹⁵ The result presented in Table 5.6 indicates that the estimated linear relationship between impacts and month of program operation is not statistically significant, although the estimate itself is positive (effect size = 0.07 and p-value = 0.351).

Finally, the middle panel of Table 5.6 indicates that impacts on outcomes other than reading comprehension across the subgroups of sites based on ERO program implementation duration during the first year are not statistically significant.

Impacts Associated with a Combination of Challenges Relating to Implementation Fidelity and Program Duration

The analysis presented in this final section of the chapter attempts to shed light on the degree to which impacts may have been stronger in schools where the challenges associated with the combination of the implementation dimensions were less serious than in schools where implementation fidelity was poorly aligned with the program models or start-up was delayed by more than six weeks. As noted in Chapter 3, many of the challenges associated with implementation fidelity and delayed start-up that were present in the first year of the project have been addressed in the second year.

As discussed in Chapter 3, the implementation of the ERO programs in 10 of the high schools was classified as poorly aligned with their program models. Also, Chapter 4 discusses the fact that 12 of the high schools experienced delays of more than six weeks in the start of their programs as they struggled to recruit and enroll students in the ERO classes and the study sample. The implementation of the ERO programs in three of these 12 schools was also classified as poorly aligned with their program models. In all, therefore, the first-year implementation experiences of 19 of the 34 participating high schools can be seen as especially problematic, either because of inadequate implementation fidelity or because of particularly long delays in enrolling students in their ERO classes and the study sample.¹⁶

The bottom panel of Table 5.6 provides a summary of impacts for schools that were able both to reach at least a moderately aligned level of implementation fidelity and to operate for

¹⁵A value ranging from six months to eight and a half months was calculated for each school. The interaction between this indicator and the treatment indicator was added to the impact estimation model. The parameter estimate for this interaction term indicates whether the ERO program impact increased or decreased as a linear function of the length of time that the programs were in operation.

¹⁶This includes (1) seven high schools that experienced poorly aligned implementation, even though they were able to begin the classes within six weeks of the start of the school year; (2) nine high schools that experienced a start-up delay of more than six weeks, even though the implementation of their ERO programs ended up being classified as at least moderately aligned with their program models; and (3) three high schools that experienced both poorly aligned implementation and a start-up delay of more than six weeks.

more than seven and a half months during the first year of the study. The ERO programs in these 15 high schools reflect conditions that were closer to those intended by the design of the demonstration than in the remaining 19 high schools that did not meet one or both of these conditions.

The bottom panel of Table 5.6 shows first that the ERO programs produced positive and statistically significant impacts on reading comprehension test scores in the 15 schools where the ERO programs were classified as at least moderately aligned with the program model and began operation within six weeks of the start of the school year. The difference between the impacts on reading comprehension for these schools and for the remaining 19 schools is an effect size of 0.16 standard deviation. This difference in impacts is statistically significant and is consistent with the hypothesis that a combination of higher-fidelity implementation and a more timely start-up (longer duration) may contribute positively to stronger impacts on reading comprehension.

Conclusion

The early impact findings indicate that, overall, the literacy programs in the ERO study produced a statistically significant improvement in students' reading comprehension skills during the first year of implementation. The findings for the ERO programs that experienced a stronger start-up provide an indication of the effectiveness of the supplemental literacy programs under conditions more reflective of the intent of the ERO project. These conditions include implementation fidelity that was moderately aligned with the ERO program model and an operating period that was more than seven and a half months. In the schools where both of these conditions were in place, the ERO programs produced a larger impact on the reading comprehension skills of struggling adolescent readers.

Although the ERO programs produced some improvement in reading comprehension test scores, students in the ERO group continued to lag behind the average ninth-grade student nationally. The 90.1 average standard score achieved by students in the ERO group at the end of their ninth-grade year corresponds, approximately, to the 6.1 grade equivalent and the 25th percentile nationally. Even when schools that experienced the most significant challenges with the first-year implementation are excluded, the more substantial impact on reading comprehension test scores for the remaining schools still left many students well below grade level. In fact, almost 90 percent of the students in both the ERO and the non-ERO group were still reading below grade level at the end of their ninth-grade year, and 76 percent of the students in the ERO group were two or more years below grade level and, thus, would still be eligible for the ERO programs, as specified by the criteria used for this project.

The early impact findings discussed in this report do not represent conclusive evidence about the efficacy or effectiveness of the supplemental literacy interventions being tested. Recognizing the need for the participating schools and teachers to gain more experience with the

programs, the U.S. Department of Education built into the design of the ERO project a second year of implementation and a second cohort of ninth-grade students for the study sample. The next report from the ERO study will provide evidence on the impact of the supplemental literacy programs during this second year of implementation. A critical goal of the second year of the study is for the participating schools and teachers to address the start-up challenges that arose in the first year and to apply their experiences from the first year and subsequent additional training.

As of this writing, the ERO study has begun to examine implementation data from the second year of the study. Twenty-seven of the 34 teachers who taught the ERO classes in the first year of the study returned for the second year. These teachers and the seven replacement teachers participated in a summer training institute and continued to learn more about how to use the instructional strategies that lie at the heart of the two interventions. All these teachers remained with their ERO programs throughout the second year. A second cohort of ninth-grade students was identified for the 2006-2007 school year. Across the 34 schools, ERO classes began within an average of approximately two weeks of the start of the school year and, at 18 of the schools, began on the first day of school.

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, subsequent reports from the evaluation will examine the impact of the programs on student performance in their core academic classes, their grade-to-grade promotion rates, and their performance on high-stakes tests required by their states. The final report from the project will present impacts on these outcomes through the eleventh grade for students in the study's first cohort and through the tenth grade for students in the second cohort.

Appendix A

ERO Student Follow-Up Survey Measures

Two surveys were administered during the first year of the ERO study. The Student Background Questionnaire, completed by all the student participants early in the 2005-2006 school year, included questions to ensure that random assignment was effective in dividing students evenly between the ERO and non-ERO groups.

This appendix describes the development of measures created from the ERO Student Follow-Up Survey. The survey was administered to students in the study near the end of their ninth-grade year, during the spring of 2006. The questions on this survey were intended to assess whether students participated in literacy support activities during the school year and to measure student attitudes and behaviors related to high school, in general, and to reading activities, in particular. A variety of measures were constructed by combining conceptually and empirically linked items from the survey. The ERO study team used a three-step process for defining and constructing the measures discussed in this appendix:

- Identify groups of conceptually linked survey items
- Conduct empirical tests of the correlation among the conceptually linked survey items
- Construct multi-item outcome variables that combine the most highly correlated items

A copy of the ERO Student Follow-up Survey is included at the end of this appendix.

Measures of Self-Reported Participation in Supplemental Literacy Support Activities

This section of the appendix describes four measures that assess the duration and frequency of student participation in supplemental literacy support activities: (1) attending a reading or writing class that took place in school; (2) working with a reading or writing tutor in school; (3) attending a reading or writing class that took place outside school; and (4) working with a reading or writing tutor outside school. Questions about the first of these activities were intended to determine whether students identified themselves as being enrolled in the ERO classes or similar types of classes that may have been offered in their high schools. Student reports about their participation in the other three activities were intended to provide an indication of the extent to which they utilized supplemental literacy support activities outside the ERO classes or similar classes that may have been offered in the participating high schools. The overall contrast between the ERO and non-ERO groups on these measures provides an indication of whether the ERO programs added literacy support activities to the landscape of what would have been available to students without the programs, at least as reported by the students in the study sample.

Each of the four measures was created based on three survey items. The first item (questions 9, 12, 15, and 18) asks whether or not a student received any of these variations of extra help. (The response choices were “Yes” or “No.”) The second item (questions 10, 13, 16, and 19) asks about the duration of this support. The response choices were on the following scale for the duration item:

- 1 = “One month”
- 2 = “A couple of months”
- 3 = “One semester or term”
- 4 = “Most of the year”
- 5 = “All year”

The third item (questions 11, 14, 17, and 20) asks about the frequency of this support. The response choices were on the following scale for the frequency item:

- 1 = “Less than once a month”
- 2 = “Once a month”
- 3 = “Every other week”
- 4 = “Once a week”
- 5 = “Twice a week”
- 6 = “3-4 times a week”
- 7 = “Every day”

Combined responses to these three items were used to construct a measure of the total number of times during the school year that a student participated in each of the four activities. If a student answered “No” to questions 9, 12, 15 or 18, the participation measure for the activity was coded to zero (0). For students who answered “Yes” to questions 9, 12, 15 or 18, Appendix Table A.1 lists the participation values calculated for every combination of answers to the questions about duration and frequency. The columns represent duration, “how long” a student received extra help (questions 10, 13, 16, and 19). The rows represent frequency, “how often” a student received that help (questions 11, 14, 17, 20). Duration and frequency were multiplied to create a measure of total participation throughout the school year for each student. The calculations are based on the assumption that there are 36 weeks of classes per school year and five days of classes per week.

Measures of Self-Reported Reading Behaviors

The ERO Student Follow-Up Survey included 29 items aimed at measuring the frequency with which students read various texts. The ERO study team developed separate measures for reading that was related to school and reading that was not related to school. In selecting items for these two measures, the team focused on the questions about written text that were

The Enhanced Reading Opportunities Study

Appendix Table A.1

Intensity Values for Supplemental Literacy Support Measures

	One month (4 weeks)	A couple of months (8 weeks)	One semester or term (18 weeks)	Most of the year (27 weeks)	All year (36 weeks)
Less than once a month (*0.1)	0.4	0.8	1.8	2.7	3.6
Once a month (*0.25)	1	2	4.5	6.75	9
Every other week (*0.5)	2	4	9	13.5	18
Once a week (*1)	4	8	18	27	36
Twice a week (*2)	8	16	36	54	72
Three to four times a week (*3.5)	14	28	63	94.5	126
Every day (*5)	20	40	90	135	180

likely to include extended passages. There was also a focus on groups of items for which student responses were highly correlated (that is, groups that were correlated with Cronbach's alpha $\geq .70$). The seven items used to construct a measure of in-school reading frequency were correlated with Cronbach's alpha = .83 and the seven measures used to construct a measure of out-of-school reading were correlated with Cronbach's alpha = .73.

The study team also developed a measure of the frequency with which a student used two reading strategies that may be characterized as "reflective" in that students would be expected to pause and think about what they were reading in order to enhance their understanding. These are strategies used by proficient readers and ones that are incorporated into the instruction of the two supplemental literacy programs for this study.¹

¹Biancarosa and Snow (2004).

Frequency of In-School Reading (7 items, Cronbach's alpha = .83)

This construct is designed to measure the frequency with which students read extended texts for school, both during the school day and for homework. It combines student responses to questions about how often they read seven types of text during the previous month. Each possible answer is converted into a value based on the approximate number of sessions the student reported reading these materials during the past month. The values for each of the seven types of texts were summed. If a student did not respond to an item, the value for that item was imputed using the mean of the values for the other items. If more than three of the items were missing, the entire construct was coded as missing for a given student.

Question 22. The items below are things you may have read for your English and other classes this year, both in class and for homework. Please indicate about how OFTEN, during the past month, you READ each of the following.

- a. History textbook
- b. Science textbook
- c. Math textbook
- d. Novels, short stories, plays, poetry or essays
- e. Research papers, reports, graphs, charts or tables
- g. Newspaper or magazine articles
- k. Workbook

Scale:

- 1 = "Never" = 0 sessions counted for the category
- 2 = "At least once" = 1 session
- 3 = "Every other week" = 2 sessions
- 4 = "Once a week" = 4 sessions
- 5 = "Twice a week" = 8 sessions
- 6 = "3-4 times a week" = 15 sessions
- 7 = "Every day" = 30 sessions

Frequency of Out-of-School Reading (7 items, Cronbach's alpha = .73)

This construct is designed to measure the frequency with which students read extended texts outside school. It combines student responses to questions about how often they read seven types of text during the previous month. Each possible answer is converted into a value based on the approximate number of sessions the student reported reading a given type of material during the past month. The values for each of the seven types of texts were summed. If a student did not respond to an item, the value for that item is imputed using the mean of the val-

ues for the other items. If more than four of the items were missing, the entire construct was coded as missing.

*Question 5. During the past month, about how *OFTEN* did you *READ* each of the following when you were not in school?*

- b. Fiction books or stories
- d. Poetry
- e. Biographies or autobiographies
- f. Books about science
- h. Books about history
- i. Newspaper or magazine articles
- k. Religious books

Scale:

- 1 = “Never” = 0 sessions counted for the category
- 2 = “At least once” = 1 session
- 3 = “Every other week” = 2 sessions
- 4 = “Once a week” = 4 sessions
- 5 = “Twice a week” = 8 sessions
- 6 = “3-4 times a week” = 15 sessions
- 7 = “Every day” = 30 sessions

Use of Reflective Reading Strategies (8 items, Cronbach’s alpha = .88)

This construct attempts to measure the degree to which students use reading strategies in which they reflect on what they are reading and ask questions of the text to better understand what they read. These measures both are consistent with the strategies taught by the ERO programs and are seen as antecedents to reading proficiency. The survey items were asked in the context of the reading that students do for English/language arts, science, history, and math classes. Since a number of students in the study sample were not taking all of these classes and did not answer all of the questions, the construct is created by averaging student responses to the first two subjects with nonmissing items in the order that the subjects are listed above.

*Question 23. Please indicate how much you *DISAGREE* or *AGREE* with the following statements about your *English* class.*

- a. I ask myself questions to make sure I know the material that I have been studying for English class.
- e. When I’m reading for English class I stop once in a while and go over what I have read.

Question 24. Please indicate how much you DISAGREE or AGREE with the following statements about your math class.

- a. I ask myself questions to make sure I know the material that I have been studying for math class.
- e. When I'm reading for math class I stop once in a while and go over what I have read.

Question 26. Please indicate how much you DISAGREE or AGREE with the following statements about your science class.

- a. I ask myself questions to make sure I know the material that I have been studying for science class.
- e. When I'm reading for science class I stop once in a while and go over what I have read.

Question 28. Please indicate how much you DISAGREE or AGREE with the following statements about your history class.

- a. I ask myself questions to make sure I know the material that I have been studying for history class.
- e. When I'm reading for history class I stop once in a while and go over what I have read.

Scale: 1 = "Strongly disagree" to 4 = "Strongly agree"

Other Measures of Student Attitudes, Perceptions, and Behaviors

The study team developed several other measures to assess the impact of the ERO program on students' attitudes toward and perceptions of reading, their engagement in school, and their educational aspirations. The creation of each of these measures is described below.

Positive Literacy Attitudes (4 items, Cronbach's alpha = .76)

This construct was designed to measure student attitudes toward reading and writing. The measure reflects the average of a student's responses to the items below. If a student did not respond to at least two of the items, the measure was coded as missing.

Question 4. Please indicate how much you DISAGREE or AGREE with the statements below about reading and writing.

- a. When I read books, I learn a lot.
- b. Reading is one of my favorite activities.
- c. Writing things like stories or letters is one of my favorite activities.
- d. Writing helps me share my ideas.

Scale: 1 = “Strongly disagree” to 4 = “Strongly agree”

Reading to Learn (3 items, Cronbach’s alpha = .74)

This construct was designed to measure how strongly a student connects reading with learning new things. It was created by averaging student responses to the items below. If a student did not respond to at least two items, the measure was coded as missing.

Question 4. Please indicate how much you DISAGREE or AGREE with the statements below about reading and writing.

- a. When I read books, I learn a lot.
- g. I read to see what is going on in the world, the country, and/or my community.
- i. I read in order to learn new things.

Scale: 1 = “Strongly disagree” to 4 = “Strongly agree”

Ease of Reading (7 items, Cronbach’s alpha = .83)

This construct was designed to measure the level of difficulty that students reported regarding the reading they did for school. It was created by averaging student responses to questions about how easy it is to read seven types of texts. If a student did not respond to at least four of these items, the construct was coded as missing.

Question 21. The statements below are about things you may have read for your English and other classes this year, both in class and for homework. Please indicate how much you DISAGREE or AGREE with each statement.

- a. My history textbook is easy to read.
- b. My science textbook is easy to read.
- c. My math textbook is easy to read.
- d. Novels, short stories, plays, poetry, or essays are easy to read.
- e. Research papers, reports, graphs, charts, or tables are easy to read.
- g. Newspaper or magazine articles are easy to read.
- k. Workbooks are easy to read.

Scale: 1 = “Strongly disagree” to 4 = “Strongly agree”

Persistence on School Work (8 items, Cronbach's alpha = .87)

This construct attempts to measure a student's persistence in completing school work. The survey items were asked in the context of the work students do for English/language arts, science, history, and math classes. Because a sizable number of students in the study sample were not taking all of these classes and did not answer all of the questions, the measure was created by averaging student responses to the first two subjects with nonmissing items, in the order suggested above.

Question 23. Please indicate how much you DISAGREE or AGREE with the following statements about your English class.

- c. Even when English study materials are dull and uninteresting, I keep working until I finish.
- f. I work hard to learn even when I don't like my English class.

Question 24. Please indicate how much you DISAGREE or AGREE with the following statements about your math class.

- c. Even when math study materials are dull and uninteresting, I keep working until I finish.
- f. I work hard to learn even when I don't like my math class.

Question 26. Please indicate how much you DISAGREE or AGREE with the following statements about your math class.

- c. Even when science study materials are dull and uninteresting, I keep working until I finish.
- f. I work hard to learn even when I don't like my science class.

Question 28. Please indicate how much you DISAGREE or AGREE with the following statements about your math class.

- c. Even when history study materials are dull and uninteresting, I keep working until I finish.
- f. I work hard to learn even when I don't like my history class.

Scale: 1 = "Strongly disagree" to 4 = "Strongly agree"

Negative School Behavior (4 items, Cronbach's alpha = .71)

This construct attempts to measure whether or not a student reported engaging in repeated negative behaviors in school during the semester. Using the four sections of Question 2 (shown below), four binary variables were created and then added together to create a cumulative variable (0-4) that suggests the level of a student's misbehavior in school. These binary variables are coded as "1" if: the student reported being late for school at least 7-9 times; the student reported that he/she cut classes at least 3-6 times; the students reported that he/she got into trouble for not following school rules at least 3-6 times; or the student reported that he/she was suspended or put on probation at least 1-2 times. If a student did not answer at least two of the items, the measure was coded as missing.

Question 2. How many times did the following things happen to you this semester or term of this school year?

- a. I was late for school.
- b. I cut or skipped classes.
- c. I got in trouble for not following school rules.
- d. I was suspended or put on probation.

Scale:

- 1 = "Never"
- 2 = "1-2 times"
- 3 = "3-6 times"
- 4 = "7-9 times"
- 5 = "10 or more times"

Educational Aspirations

This question is designed to measure a student's aspirations for educational attainment. It is coded as a binary variable that equals one if the student plans to graduate from a four-year college or higher (response codes 5, 6, or 7) and zero if the student does not plan to graduate from a four-year college (response codes 1, 2, 3, or 4).

Question 3. How far do you think you will go in school?

1. Graduate from high school
2. Vocational or technical training
3. Some college
4. Graduate from a business or two-year college
5. Graduate from a four-year college
6. Get a master's degree
7. Get a law degree, a Ph.D., or a medical doctor's degree



**STUDENT FOLLOW-UP QUESTIONNAIRE
SPRING 2006
GRADE 9**

First Name: «First Name»

Last Name: «Last Name»

School: «School»

Student ID #: «Student_ID_Number»

Date of Birth: «Month»/«Day»/«Year»
Month Day Year

Today's Date: _____ / _____ / _____
Month

Day Year

PURPOSE

We are asking you these questions to get information about your school experiences and your experiences with reading. You're the best person to help us learn about these things. We are interested in your own responses to these questions. You do not need to ask your parents, teachers, or friends for help on the answers.

This is not a test – there are no right or wrong answers. Your answers will be used for research only, so please be as honest as you can.

You do not have to answer any individual questions you don't like. We hope that you answer all the questions because we need your answers to make our research complete.

DIRECTIONS

Read each question carefully. Try to answer all questions. If no answer fits exactly, pick the one that comes closest. It is important that you follow the directions for responding to each question. Mark (✓) each answer clearly.

YOUR ANSWERS WILL BE USED FOR RESEARCH ONLY.

MDRC, New York, NY, www.mdrc.org

For questions, contact Jim Kemple at: James.Kemple@mdrc.org, Phone: (866)519-1884

The U.S. Department of Education wants to protect the privacy of individuals who participate in surveys. Your answers will be combined with other surveys, and no one will know how you answered the questions. This survey is authorized by law (1) Sections 171(b) and 173 of the Education Sciences Reform Act of 2002, Pub. L. 107-279 (2002); and (2) Section 9601 of the Elementary and Secondary Education Act (ESEA), as amended by the No Child Left Behind (NCLB) Act of 2001 (Pub. L. 107-110).

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. The valid OMB control number for this information collection is **1850-0801**. The time required to complete this information collection is estimated to be 25 minutes per respondent, including the time to review instructions, respond to the questions, and review the responses. If you have any comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to: U.S. Department of Education, Washington, DC 20202. If you have comments or concerns regarding the status of your individual submission of this form, write directly to: U.S. Department of Education, Institute of Education Sciences, 555 New Jersey Avenue, NW, Washington, DC 20208.

FOR SURVEY ADMINISTRATOR USE ONLY

Non-ERO School Administration

First, we have two general questions about going to school.

Mark (✓) the number on each line that applies to you.

(1) How much do you agree or disagree with the following statements about why you go to school?

	Strongly Disagree	Disagree	Agree	Strongly Agree
a. I go to school because I think the subjects I'm taking are interesting and challenging.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
b. I go to school because I get a feeling of satisfaction from doing what I'm supposed to do in class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
c. I go to school because I have nothing better to do.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
d. I go to school because education is important for getting a job later on.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
e. I go to school because it's a place to meet my friends.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
f. I go to school because I play on a team or belong to a club.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
g. I go to school because I'm learning skills that I will need for a job.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
h. I go to school because my teachers expect me to succeed.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
i. I go to school because my parents expect me to succeed.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>

(2) How many times did the following things happen to you this semester or term of this school year?

	Never	1-2 Times	3-6 Times	7-9 Times	10 or More
a. I was late for school.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>
b. I cut or skipped classes.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>
c. I got in trouble for not following school rules.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>
d. I was suspended or put on probation.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>

The next question asks you about your future education.

(3) How far do you think you will go in school?

Mark (✓) one answer.

- ₁ graduate from high school
- ₂ vocational or technical training (e.g. electrician, hairdresser, chef, pre-school teacher)
- ₃ some college
- ₄ graduate from a business or two-year college
- ₅ graduate from a four-year college
- ₆ get a master's degree
- ₇ get a law degree, a Ph.D., or a medical doctor's degree

This section is about reading and writing. The section has 19 questions.

Please mark (✓) one answer on each line.

(4) Please indicate how much you DISAGREE or AGREE with the statements below about reading and writing.

	Strongly Disagree	Disagree	Agree	Strongly Agree
a. When I read books, I learn a lot.	<input type="radio"/> ₁	<input type="radio"/> ₂	<input type="radio"/> ₃	<input type="radio"/> ₄
b. Reading is one of my favorite activities.	<input type="radio"/> ₁	<input type="radio"/> ₂	<input type="radio"/> ₃	<input type="radio"/> ₄
c. Writing things like stories or letters is one of my favorite activities.	<input type="radio"/> ₁	<input type="radio"/> ₂	<input type="radio"/> ₃	<input type="radio"/> ₄
d. Writing helps me share my ideas.	<input type="radio"/> ₁	<input type="radio"/> ₂	<input type="radio"/> ₃	<input type="radio"/> ₄
e. I read or write to get away from family or friends.	<input type="radio"/> ₁	<input type="radio"/> ₂	<input type="radio"/> ₃	<input type="radio"/> ₄
f. I read or write when there's no one else to talk or be with.	<input type="radio"/> ₁	<input type="radio"/> ₂	<input type="radio"/> ₃	<input type="radio"/> ₄
g. I read to see what is going on in the world, the country, and/or my community.	<input type="radio"/> ₁	<input type="radio"/> ₂	<input type="radio"/> ₃	<input type="radio"/> ₄
h. I read or write when I have nothing better to do or when I am bored.	<input type="radio"/> ₁	<input type="radio"/> ₂	<input type="radio"/> ₃	<input type="radio"/> ₄
i. I read in order to learn new things.	<input type="radio"/> ₁	<input type="radio"/> ₂	<input type="radio"/> ₃	<input type="radio"/> ₄
j. I read or write because it's a habit, just something I do.	<input type="radio"/> ₁	<input type="radio"/> ₂	<input type="radio"/> ₃	<input type="radio"/> ₄
k. I read or write so I can forget about school, work, or other things.	<input type="radio"/> ₁	<input type="radio"/> ₂	<input type="radio"/> ₃	<input type="radio"/> ₄
l. I read or write because it makes me feel less lonely.	<input type="radio"/> ₁	<input type="radio"/> ₂	<input type="radio"/> ₃	<input type="radio"/> ₄

(5) During the past month, about how **OFTEN** did you **READ** each of the following, when you were not in school?

	Never	At least once	Every other week	Once a week	Twice a week	3-4 times a week	Every day
a. Comic books or joke books	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
b. Fiction books or stories (books or stories about imagined events)	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
c. Plays	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
d. Poetry	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
e. Biographies or autobiographies	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
f. Books about science (for example, nature, animals, astronomy)	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
g. Books about technology (for example, machines, computers)	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
h. Books about history	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
i. Newspaper or magazine articles	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
j. E-mails, letters, or notes	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
k. Religious books (e.g., Koran, Bible, Catechism, Torah, other)	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
l. Websites on the Internet	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
m. Music lyrics (words to music)	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
n. Research papers, reports, graphs, charts, or tables	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
o. Instruction manuals, cookbooks, sewing patterns (instructions on how to do something)	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
p. Maps or bus, airline, or train schedules	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
q. Catalogs or reference books (encyclopedia, dictionary, phone book, etc.)	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>

(6) During the past month, how **OFTEN** did you **READ** for fun?

	Never	At least once	Every other week	Once a week	Twice a week	3-4 times a week	Every day
	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>

(7) During the past month, about how **OFTEN** did you **WRITE** each of the following, when you were **not** in school?

	Never	At least once	Every other week	Once a week	Twice a week	3-4 times a week	Every day
a. E-mails, chat, shout-outs, blogs	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
b. A private diary or journal	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
c. Letters or notes on paper	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
d. Poetry	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
e. Stories	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
f. Grocery/shopping list	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
g. Instructions on how to do something	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
h. Music lyrics (words to music)	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
i. Directions on how to get somewhere	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
j. Graffiti or tagging on paper	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
k. Comics	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>

(8) During the past month, how **OFTEN** did you **WRITE** for fun?

Never	At least once	Every other week	Once a week	Twice a week	3-4 times a week	Every day
1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>

(9) Other than your regular English class, have you taken a class, in school this year intended to help you with your reading and writing?

Yes	No
1 <input type="radio"/> If YES, please continue to question 10	2 <input type="radio"/> If NO, please continue to question 12

(10) For how **LONG** did you get this help with reading and writing?

One month or less	A couple of months	One semester or term	Most of the year	All year
1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>

(11) How **OFTEN** did you get this help with reading and writing?

Less than once a month	Once a month	Every other week	Once a week	Twice a week	3-4 times a week	Every day
1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>

(12) Did an adult <u>in your school</u> help you individually with your reading and writing this year, like a tutor?	Yes	No
	<input type="radio"/> 1 If YES, please continue to question 13	<input type="radio"/> 2 If NO, please continue to question 15

(13) For how LONG did you get this help with reading and writing?	One month or less	A couple of months	One semester or term	Most of the year	All year
	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5

(14) How OFTEN did you get this help with reading and writing?	Less than once a month	Once a month	Every other week	Once a week	Twice a week	3-4 times a week	Every day
	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7

(15) Have you taken a class or participated in a program <u>outside of school</u> intended to help you with your reading and writing?	Yes	No
	<input type="radio"/> 1 If YES, please continue to question 16	<input type="radio"/> 2 If NO, please continue to question 18

(16) For how LONG did you get this help with reading and writing?	One month or less	A couple of months	One semester or term	Most of the year	All year
	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5

(17) How OFTEN did you get this help with reading and writing?	Less than once a month	Once a month	Every other week	Once a week	Twice a week	3-4 times a week	Every day
	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7

(18) Did an adult <u>outside of school</u> help you individually with your reading and writing this year, like a tutor or someone at an after-school program?	Yes	No
	1 <input type="radio"/> If YES, please continue to question 19	2 <input type="radio"/> If NO, please continue to question 21

(19) For how LONG did you get this help with reading and writing?	One month or less	A couple of months	One semester or term	Most of the year	All year
	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>

(20) How OFTEN did you get this help with reading and writing?	Less than once a month	Once a month	Every other week	Once a week	Twice a week	3-4 times a week	Every day
	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>

The next two questions ask about what you read in school.

(21) The statements below are about things you may have read for your English and other classes this year, both in class and for homework. Please indicate how much you DISAGREE or AGREE with each statement. Mark (✓) the number on each line that applies to you.

	Didn't read	Strongly Disagree	Disagree	Agree	Strongly Agree
a. My history textbook is easy to read.	9 <input type="radio"/>	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
b. My science textbook is easy to read.	9 <input type="radio"/>	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
c. My math textbook is easy to read.	9 <input type="radio"/>	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
d. Novels, short stories, plays, poetry, or essays are easy to read.	9 <input type="radio"/>	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
e. Research papers, reports, graphs, charts, or tables are easy to read.	9 <input type="radio"/>	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
f. Class notes are easy to read.	9 <input type="radio"/>	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
g. Newspaper or magazine articles are easy to read.	9 <input type="radio"/>	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
h. Websites on the Internet are easy to read.	9 <input type="radio"/>	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
i. Maps are easy to read.	9 <input type="radio"/>	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
j. Vocabulary lists are easy to read.	9 <input type="radio"/>	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
k. Workbooks are easy to read.	9 <input type="radio"/>	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>

(22) The items below are things you may have read for your English and other classes this year, both in class and for homework. Please indicate about how OFTEN, during the past month, you READ each of the following. Mark (✓) the number on each line that applies to you.

	Never	At least once	Every other week	Once a week	Twice a week	3-4 times a week	Every day
a. History textbook	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
b. Science textbook	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
c. Math textbook	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
d. Novels, short stories, plays, poetry, or essays	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
e. Research papers, reports, graphs, charts, or tables	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
f. Class notes	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
g. Newspaper or magazine articles	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
h. Websites on the Internet	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
i. Maps	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
j. Vocabulary lists	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>
k. Workbooks	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>

This section is about your classes in school this year. This section has 6 questions.

(23) Please indicate how much you DISAGREE or AGREE with the following statements about your English class. Mark (✓) the number on each line that applies to you.

	Strongly Disagree	Disagree	Agree	Strongly Agree
a. I ask myself questions to make sure I know the material that I have been studying for <u>English</u> class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
b. When work in <u>English</u> class is hard I either give up or study only the easy parts.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
c. Even when <u>English</u> study materials are dull and uninteresting, I keep working until I finish.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
d. I often find that I have been reading for <u>English</u> class but don't know what it is all about.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
e. When I'm reading for <u>English</u> class I stop once in a while and go over what I have read.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
f. I work hard to learn even when I don't like my <u>English</u> class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
g. I have to read well to do well in <u>English</u> class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
h. My <u>English</u> teacher teaches us things in class to help us read better.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>

(24) Please indicate how much you DISAGREE or AGREE with the following statements about your math class. Mark (✓) the number on each line that applies to you.

	Strongly Disagree	Disagree	Agree	Strongly Agree
a. I ask myself questions to make sure I know the material that I have been studying for <u>math</u> class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
b. When work in <u>math</u> class is hard I either give up or study only the easy parts.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
c. Even when <u>math</u> study materials are dull and uninteresting, I keep working until I finish.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
d. I often find that I have been reading for <u>math</u> class but don't know what it is all about.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
e. When I'm reading for <u>math</u> class I stop once in a while and go over what I have read.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
f. I work hard to learn even when I don't like my <u>math</u> class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
g. I have to read well to do well in <u>math</u> class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
h. My <u>math</u> teacher teaches us things in class to help us read better.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>

	Yes	No
(25) Did you take Science this year?	1 <input type="radio"/>	2 <input type="radio"/>
	If YES, please continue to question 26	If NO, please continue to question 27

(26) Please indicate how much you DISAGREE or AGREE with the following statements about your science class. Mark (✓) the number on each line that applies to you.

	Strongly Disagree	Disagree	Agree	Strongly Agree
a. I ask myself questions to make sure I know the material that I have been studying for <u>science</u> class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
b. When work in <u>science</u> class is hard I either give up or study only the easy parts.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
c. Even when <u>science</u> study materials are dull and uninteresting, I keep working until I finish.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
d. I often find that I have been reading for <u>science</u> class but don't know what it is all about.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
e. When I'm reading for <u>science</u> class I stop once in a while and go over what I have read.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
f. I work hard to learn even when I don't like my <u>science</u> class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
g. I have to read well to do well in <u>science</u> class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
h. My <u>science</u> teacher teaches us things in class to help us read better.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>

	Yes	No
(27) Did you take History (or social studies) this year?	1 <input type="radio"/>	2 <input type="radio"/>
	If YES, please continue to question 28	If NO, please continue to question 29

(28) Please indicate how much you DISAGREE or AGREE with the following statements about your history class. Mark (✓) the number on each line that applies to you.

	Strongly Disagree	Disagree	Agree	Strongly Agree
a. I ask myself questions to make sure I know the material that I have been studying for <u>history</u> class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
b. When work in <u>history</u> class is hard I either give up or study only the easy parts.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
c. Even when <u>history</u> study materials are dull and uninteresting, I keep working until I finish.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
d. I often find that I have been reading for <u>history</u> class but don't know what it is all about.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
e. When I'm reading for <u>history</u> class I stop once in a while and go over what I have read.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
f. I work hard to learn even when I don't like my <u>history</u> class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
g. I have to read well to do well in <u>history</u> class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
h. My <u>history</u> teacher teaches us things in class to help us read better.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>

This final section is about your Enhanced Reading Opportunity (ERO) class (Xtreme Reading or Reading Apprenticeship For Academic Literacy). There are 3 questions.

(29) Please indicate how much you DISAGREE or AGREE with the following statements about your ERO class. Mark (✓) the number on each line that applies to you.

	Strongly Disagree	Disagree	Agree	Strongly Agree
a. I like my ERO class.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
b. Compared to work I do for other subjects at school, I find the work I do for ERO to be interesting.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>
c. Compared with what I learn in my other subjects at school, I find what I learn in ERO to be useful.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>

THANK YOU!!!

Appendix B

Follow-Up Test and Survey Response Analysis

The two main data sources for the first-year impact analysis of the ERO program are the GRADE assessment of student reading skills and the student follow-up survey. Both the test and the survey were administered late in the 2005-2006 school year. Approximately 83 percent of the full study sample completed the test and survey, including 84 percent of students in the ERO program group and 81 percent of students in the non-ERO group. The lack of a 100 percent response rate combined with the discrepancy between response rates for the ERO and non-ERO student groups raises two concerns: Are the respondents representative of the full study sample? Are there systematic pre-program differences between respondents in the ERO and non-ERO groups?

The first section of this appendix discusses the follow-up test and survey response rates and examines differences between respondents and nonrespondents. The second section examines the respondent sample and assesses similarities and differences between students in the ERO and non-ERO groups.

Follow-Up Test and Survey Response Rates

Efforts were made to collect both test and survey data from all 2,916 students who make up the full study sample — ninth-grade students who consented to be in the ERO program and had pretest reading comprehension scores between the fourth- and seventh-grade levels. Sections of 25 to 30 students from both the ERO and the non-ERO group were tested and surveyed together in their high schools. The test and survey administrations took place during the school day and were proctored by members of the ERO study team. The ERO study team spent up to four days at each school locating, testing, and surveying students who did not attend the originally scheduled session.

In all, 2,397 students (82 percent of the full study sample) completed both the follow-up test and the survey. An additional 16 students completed only the follow-up test, and 15 students completed only the survey. Due to the similarity in response rates for the follow-up test and the survey, the non-response analysis in this appendix focuses on the response rate for the test. Results for the survey response and the combined response are virtually the same.

Appendix Table B.1 shows the follow-up test response rates for all 34 participating high schools combined and for the groups of schools using Reading Apprenticeship Academic Literacy and Xtreme Reading, respectively. Overall, 84 percent of students in the ERO group took the follow-up test, compared with 81 percent of students in the non-ERO group. The three percentage point difference is statistically significant (p-value is less than or equal to 5 percent). The Reading Apprenticeship and Xtreme Reading schools had very similar response rates for their ERO group students. The difference in response rates between the ERO and non-ERO

groups is statistically significant for the Reading Apprenticeship schools but not for the Xtreme Reading schools.

The primary reason that students did not complete the follow-up test or survey is that they were no longer enrolled in a high school participating in the ERO study.¹ In all, approximately 10 percent of the students in the study sample were no longer enrolled in an ERO high school at the time of the follow-up test and survey administrations. These rates are similar for the ERO group (11 percent) and the non-ERO group (9 percent). Of the students who were no longer enrolled in an ERO school, only 15 percent completed the follow-up test (compared with 91 percent of those who remained enrolled in an ERO school). These completion rates were the same for students in the ERO and non-ERO groups who were no longer enrolled in an ERO school. As with the full sample, response rates for students who remained in an ERO school differ somewhat for the ERO group (93 percent) compared with the non-ERO group students (89 percent). As with the full sample, this difference was concentrated in the Reading Apprenticeship schools, where 94 percent of the ERO group completed the follow-up test, compared with 90 percent of the non-ERO group.

One factor that may influence the interpretation of the impact findings presented in this report is whether students who completed the follow-up test and survey are representative of the full study sample. This question was addressed in two ways. First, respondents and nonrespondents were compared directly on a range of background characteristics. The results for the full study sample are shown in Appendix Table B.2. Overall, the table indicates that nonrespondents were more likely than respondents to have characteristics associated with a risk of school failure. For example, nonrespondents were more likely to be overage for the ninth grade (indicating that they were likely to have been retained in a prior grade) and more likely to have a parent who did not complete high school. Also, nonrespondents had lower baseline reading comprehension test scores, on average, than students who completed the follow-up test. Appendix Tables B.3 and B.4 compare the respondents and nonrespondents in Reading Apprenticeship schools and Xtreme Reading schools, respectively.

A second and more comprehensive strategy for assessing differences between respondents and nonrespondents is to use multiple regression to determine the extent to which the average characteristics of students who completed the follow-up test differ systematically from those who did not. This analysis was carried out for the full group of schools in the study and separately for the schools using Reading Apprenticeship and Xtreme Reading, respectively. The results are presented in Appendix Table B.5. It indicates that response rates differ by high

¹The tracking information on reasons that students did not complete the follow-up test or survey is based on data collected during the administration period and is available only in aggregate form. As a result, it does not permit breakdowns by student background characteristics.

school as well as by several background characteristics including overage for ninth grade, parents' education, and baseline test scores. More important, the overall F-test for each regression indicates that there are systematic differences between the respondents and nonrespondents.

In summary, the response analysis indicates that students who completed the follow-up test and survey are not fully representative of the full study sample of 2,916 students. Thus, some caution should be exercised when attempting to generalize the findings beyond those who are included in the impact analysis. Nevertheless, the overall response rates show that follow-up data are available for 83 percent of the students in the study sample, making the results reflective of the behavior of most of the targeted students.

Appendix F presents an assessment of the sensitivity of the impact findings to differences between students who completed the follow-up test and those who did not. The appendix presents estimated impacts that are weighted for differential response rates by high school, overage for grade, pretest scores, and research status. These analyses yield impact estimates that are similar to those presented in the text of the report.

Characteristics of Students Who Completed the Follow-Up Test and Survey

The random assignment research design ensures that there are no systematic differences in measured and unmeasured characteristics between the students in the sample who were assigned to the ERO group and those who were not. Because the two groups began the study with equivalent characteristics, any differences that emerge after random assignment can be attributed with confidence to the fact that one group had access to the ERO programs and the other did not.

When completion rates for follow-up data collection are less than 100 percent, a key question underlying the impact analyses is: Do the response rates preserve the random assignment design? In other words, does the sample of students who completed the follow-up test and survey exhibit the same lack of systematic differences between the ERO and non-ERO groups, both overall and for groups of sites using Reading Apprenticeship and Xtreme Reading? To address this question, multiple regression was used to assess whether there are systematic differences in background characteristics between the ERO and non-ERO groups. The results are presented in Appendix Table B.6. The overall F-tests for these regressions indicate that there are no systematic differences between the two groups, either overall or for the Reading Apprenticeship and Xtreme Reading schools. Further, none of the individual parameter estimates in the regressions are statistically significant (p -value is less than or equal to 5 percent).

Comparisons in Chapter 2 of students in the ERO and non-ERO groups are also displayed in Table 2.4 for all 34 high schools in the study, in Table 2.5 for the Reading Appren-

ticeship schools, and in Table 2.6 for the Xtreme Reading schools. Each of these tables indicates a high degree of similarity between students in the ERO and non-ERO groups.

In summary, the follow-up test and survey completion rates preserve the random assignment design for the ERO study in terms of the characteristics of students measured at baseline. As a result, one may have a high degree of confidence that any differences found in the follow-up data reflect the impact of the ERO programs.

The Enhanced Reading Opportunities Study
Appendix Table B.1
Response Rates of Students in Cohort 1
Full Study Sample

	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
All schools	84.1	81.1	2.9 *	0.037
Reading Apprenticeship schools	84.6	79.3	5.2 *	0.011
Xtreme Reading schools	83.6	82.7	0.9	0.649
Strong start-up schools	83.1	79.4	3.8	0.067
Weak start-up schools	84.9	82.8	2.1	0.268
Overage for grade ^a	75.0	70.0	5.0	0.104
Not overage for grade	88.2	85.7	2.5	0.093
Language other than English spoken at home	86.9	82.7	4.2 *	0.033
English only spoken at home	81.7	80.2	1.5	0.442
Baseline reading comprehension score				
6.0 - 7.0 grade equivalent	87.2	83.0	4.2	0.062
5.0 - 5.9 grade equivalent	83.4	76.5	7.0 *	0.010
4.0 - 4.9 grade equivalent	81.7	82.1	-0.3	0.885
Sample size	1,675	1,241		

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

NOTES: This table represents the response rates for the follow-up GRADE assessment which was administered in spring 2006 at the end of students' ninth-grade year. The follow-up student questionnaire was also administered at that time. The difference in response rates between the test and survey is negligible.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. The p-value is the probability that the observed difference is the result of chance and does not represent a true difference between groups. The lower the p-value, the less confidence that there is not a difference between the two groups. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

The Enhanced Reading Opportunities Study

Appendix Table B.2

Characteristics of Students in Cohort 1:
Differences Between Respondents and Nonrespondents

Characteristic	Respondents	Non- Respondents	Difference	P-Value for the Difference
Race/ethnicity (%)				
Hispanic	33.3	33.5	-0.2	0.907
Black, non-Hispanic	43.1	45.5	-2.4	0.189
White, non-Hispanic	17.7	16.6	1.1	0.503
Other	6.0	4.5	1.5	0.182
Gender (%)				
Male	50.8	47.2	3.6	0.151
Female	49.2	52.8	-3.6	0.151
Average age (years)	14.7	15.1	-0.3 *	0.000
Overage for grade ^a (%)	26.7	45.7	-19.0 *	0.000
Language other than English spoken at home (%)	47.0	44.1	2.8	0.202
Language spoken at home missing (%)	6.8	6.6	0.3	0.809
Mother's education level (%)				
Did not finish high school	17.0	27.7	-10.7 *	0.000
High school diploma or GED certificate	25.2	22.8	2.4	0.258
Completed some postsecondary education	29.9	26.5	3.3	0.129
Don't know	20.4	16.2	4.2 *	0.032
Missing	7.5	6.8	0.7	0.530
Father's education level (%)				
Did not finish high school	16.9	21.2	-4.4 *	0.019
High school diploma or GED certificate	22.7	24.2	-1.5	0.477
Completed some postsecondary education	19.8	15.8	4.0 *	0.038
Don't know	32.2	30.0	2.2	0.340
Missing	8.4	8.8	-0.4	0.780
GRADE reading comprehension ^b				
Average standard score	86.0	85.3	0.7 *	0.004
<i>Corresponding grade equivalent</i>	<i>5.1</i>	<i>5.0</i>		
<i>Corresponding percentile</i>	<i>16</i>	<i>15</i>		
6.0 - 7.0 grade equivalent (%)	35.4	28.7	6.8 *	0.004
5.0 - 5.9 grade equivalent (%)	28.2	29.9	-1.7	0.445
4.0 - 4.9 grade equivalent (%)	36.4	41.4	-5.1 *	0.032
Sample size	2,413	503		

(continued)

Appendix Table B.2 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year.

The differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The respondents value is the unadjusted mean for the students in the respondent sample. The non-respondents value is the respondents value minus the difference.

A two-tailed t-test was used to test differences between the respondents and non-respondents. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

^bThe 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 A). No statistical tests or arithmetic operations were performed on these reference points.

The Enhanced Reading Opportunities Study

Appendix Table B.3

Characteristics of Students in Cohort 1:
Differences Between Respondents and Nonrespondents,
Reading Apprenticeship Schools

Characteristic	Respondents	Non- Respondents	Difference	P-Value for the Difference
Race/ethnicity (%)				
Hispanic	32.5	31.4	1.2	0.606
Black, non-Hispanic	43.0	43.7	-0.7	0.786
White, non-Hispanic	18.2	19.8	-1.5	0.518
Other	6.2	5.2	1.0	0.535
Gender (%)				
Male	50.9	46.3	4.5	0.202
Female	49.1	53.7	-4.5	0.202
Average age (years)	14.7	15.0	-0.3 *	0.000
Overage for grade ^a (%)	26.1	42.9	-16.9 *	0.000
Language other than English spoken at home (%)	45.7	44.9	0.8	0.802
Language spoken at home missing (%)	7.3	7.5	-0.2	0.886
Mother's education level (%)				
Did not finish high school	17.5	24.8	-7.3 *	0.008
High school diploma or GED certificate	24.6	26.0	-1.5	0.635
Completed some postsecondary education	28.5	24.1	4.4	0.164
Don't know	21.6	17.3	4.3	0.136
Missing	7.9	7.7	0.2	0.918
Father's education level (%)				
Did not finish high school	17.0	20.7	-3.7	0.169
High school diploma or GED certificate	22.3	22.4	-0.1	0.968
Completed some postsecondary education	18.2	13.2	4.9	0.064
Don't know	33.5	32.4	1.1	0.748
Missing	9.0	11.2	-2.2	0.253
GRADE reading comprehension ^b				
Average standard score	86.0	85.0	1.1 *	0.004
<i>Corresponding grade equivalent</i>	<i>5.2</i>	<i>5.0</i>		
<i>Corresponding percentile</i>	<i>17</i>	<i>15</i>		
6.0 - 7.0 grade equivalent (%)	36.0	26.4	9.6 *	0.004
5.0 - 5.9 grade equivalent (%)	28.6	30.1	-1.5	0.633
4.0 - 4.9 grade equivalent (%)	35.4	43.5	-8.1 *	0.018
Sample size	1,140	245		

(continued)

Appendix Table B.3 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year.

The differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The respondents value is the unadjusted mean for the students in the respondent sample. The non-respondents value is the respondents value minus the difference.

A two-tailed t-test was used to test differences between the respondents and non-respondents. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

^bThe 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 A). No statistical tests or arithmetic operations were performed on these reference points.

The Enhanced Reading Opportunities Study

Appendix Table B.4

Characteristics of Students in Cohort 1:
Differences Between Respondents and Nonrespondents,
Xtreme Reading Schools

Characteristic	Respondents	Non- Respondents	Difference	P-Value for the Difference
Race/ethnicity (%)				
Hispanic	33.9	35.4	-1.5	0.494
Black, non-Hispanic	43.1	47.1	-4.0	0.111
White, non-Hispanic	17.1	13.6	3.5	0.109
Other	5.8	3.8	2.0	0.203
Gender (%)				
Male	50.7	48.0	2.7	0.447
Female	49.3	52.0	-2.7	0.447
Average age (years)	14.7	15.1	-0.4 *	0.000
Overage for grade ^a (%)	27.3	48.3	-21.0 *	0.000
Language other than English spoken at home (%)	48.1	43.3	4.8	0.121
Language spoken at home missing (%)	6.4	5.7	0.7	0.603
Mother's education level (%)				
Did not finish high school	16.6	30.4	-13.8 *	0.000
High school diploma or GED certificate	25.8	19.7	6.1 *	0.041
Completed some postsecondary education	31.1	28.7	2.4	0.444
Don't know	19.3	15.2	4.1	0.123
Missing	7.1	5.9	1.2	0.419
Father's education level (%)				
Did not finish high school	16.7	21.8	-5.0	0.052
High school diploma or GED certificate	23.1	25.9	-2.8	0.344
Completed some postsecondary education	21.3	18.2	3.1	0.265
Don't know	31.0	27.7	3.3	0.303
Missing	7.9	6.4	1.4	0.380
GRADE reading comprehension ^b				
Average standard score	86.0	85.5	0.4	0.215
<i>Corresponding grade equivalent</i>	<i>5.1</i>	<i>5.1</i>		
<i>Corresponding percentile</i>	<i>16</i>	<i>16</i>		
6.0 - 7.0 grade equivalent (%)	35.0	30.9	4.0	0.222
5.0 - 5.9 grade equivalent (%)	27.8	29.7	-1.9	0.547
4.0 - 4.9 grade equivalent (%)	37.2	39.4	-2.2	0.506
Sample size	1,273	258		

(continued)

Appendix Table B.4 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year.

The differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The respondents value is the unadjusted mean for the students in the respondent sample. The non-respondents value is the respondents value minus the difference.

A two-tailed t-test was used to test differences between the respondents and non-respondents. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

^bThe 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 A). No statistical tests or arithmetic operations were performed on these reference points.

The Enhanced Reading Opportunities Study
Appendix Table B.5
Regression Coefficients for the Probability of Being in the Respondent Sample,
Full Study Sample

Variable	Parameter Estimates (Standard Errors)		
	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Intercept	1.966 *	1.754 *	2.046 *
	(0.287)	(0.420)	(0.393)
School 1	-0.098	-0.032	
	(0.055)	(0.059)	
School 2	-0.122 *	-0.050	
	(0.058)	(0.061)	
School 3	-0.163 *		-0.168 *
	(0.056)		(0.058)
School 4	-0.099		-0.120 *
	(0.058)		(0.060)
School 5	-0.187 *		-0.186 *
	(0.054)		(0.058)
School 6	-0.247 *		-0.244 *
	(0.053)		(0.056)
School 7	-0.215 *	-0.173 *	
	(0.053)	(0.059)	
School 8	-0.056	-0.014	
	(0.054)	(0.060)	
School 9	-0.119		-0.148 *
	(0.063)		(0.066)
School 10	-0.056	0.016	
	(0.064)	(0.067)	
School 11	-0.036	0.002	
	(0.060)	(0.065)	
School 12	-0.059		-0.082
	(0.061)		(0.062)
School 13	-0.037		-0.038
	(0.057)		(0.056)
School 14	-0.055		-0.055
	(0.050)		(0.049)
School 15	-0.092	-0.052	
	(0.050)	(0.056)	
School 16	-0.151 *	-0.109	
	(0.064)	(0.069)	
School 17	-0.135 *	-0.083	
	(0.056)	(0.061)	
School 18	0.017		0.007
	(0.056)		(0.059)

(continued)

Appendix Table B.5 (continued)

Variable	Parameter Estimates (Standard Errors)		
	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
School 19	-0.096 (0.057)		-0.111 (0.059)
School 20	0.005 (0.061)	0.051 (0.067)	
School 21	-0.079 (0.052)	-0.018 (0.057)	
School 22	-0.035 (0.050)		-0.047 (0.050)
School 23	-0.047 (0.055)		-0.063 (0.057)
School 24 ^a	-0.054 (0.054)	-- --	
School 25	-0.036 (0.064)	0.010 (0.068)	
School 26	-0.052 (0.054)		-0.055 (0.057)
School 27	-0.087 (0.049)	-0.042 (0.057)	
School 28	-0.093 (0.054)	-0.052 (0.061)	
School 29	-0.026 (0.050)		-0.025 (0.049)
School 30 ^a	-- --		-- --
School 31	-0.046 (0.059)	0.008 (0.064)	
School 32	-0.020 (0.058)		-0.031 (0.059)
School 33	-0.076 (0.059)	-0.015 (0.062)	
School 34	-0.054 (0.058)		-0.073 (0.062)
Research status			
ERO group	0.033 * (0.014)	0.054 * (0.020)	0.013 (0.019)
Non-ERO group ^a	-- --	-- --	-- --

(continued)

Appendix Table B.5 (continued)

Variable	Parameter Estimates (Standard Errors)		
	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Race/ethnicity (%)			
Hispanic	-0.012 (0.028)	0.046 (0.042)	-0.060 (0.039)
Black, non-Hispanic	-0.024 (0.024)	0.026 (0.035)	-0.063 (0.032)
White, non-Hispanic ^a	--	--	--
Other	0.014 (0.034)	0.035 (0.049)	-0.006 (0.048)
Gender (%)			
Male	0.022 (0.014)	0.028 (0.021)	0.016 (0.019)
Female ^a	--	--	--
Average age (years)	-0.094 * (0.018)	-0.094 * (0.026)	-0.090 * (0.024)
Overage for grade ^b (%)	-0.024 (0.026)	-0.016 (0.039)	-0.034 (0.035)
Language other than English spoken at home (%)	0.024 (0.017)	0.006 (0.026)	0.041 (0.024)
Language spoken at home missing (%)	-0.008 (0.061)	-0.009 (0.095)	-0.008 (0.081)
Mother's education level (%)			
Did not finish high school ^a	--	--	--
High school diploma or GED certificate	0.091 * (0.023)	0.046 (0.034)	0.130 * (0.031)
Completed some postsecondary education	0.077 * (0.023)	0.050 (0.034)	0.097 * (0.031)
Don't know	0.100 * (0.025)	0.085 * (0.036)	0.106 * (0.034)
Missing	0.159 * (0.066)	0.179 (0.100)	0.113 (0.089)
Father's education level (%)			
Did not finish high school ^a	--	--	--
High school diploma or GED certificate	-0.017 (0.023)	0.002 (0.035)	-0.024 (0.031)
Completed some postsecondary education	0.010 (0.025)	0.030 (0.038)	0.001 (0.034)
Don't know	-0.004 (0.023)	-0.008 (0.033)	0.006 (0.031)
Missing	-0.068 (0.054)	-0.126 (0.074)	0.022 (0.079)

(continued)

Appendix Table B.5 (continued)

Variable	Parameter Estimates (Standard Errors)		
	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
GRADE reading comprehension Average standard score	0.003 * (0.001)	0.004 * (0.002)	0.002 (0.002)
Sample size	2,916	1,385	1,531
Degrees of freedom	51	34	34
Mean of the dependent variable	0.828	0.823	0.831
R-square	0.086	0.078	0.102
F-statistic	5.251	3.348	5.003
P-value of F-statistic	0.000	0.000	0.000

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

^aCovariates marked by '--' were not included in the regression. The site with the highest response rate was not included.

^bA student is defined as overage for grade if he or she turned 15 before the start of the ninth grade.

The Enhanced Reading Opportunities Study
Appendix Table B.6
Regression Coefficients for the Probability of Being in the Treatment Group,
Respondent Sample

Variable	Parameter Estimates (Standard Errors)		
	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Intercept	0.634 (0.448)	0.109 (0.649)	1.069 (0.622)
School 1	0.078 (0.081)	0.013 (0.085)	
School 2	0.093 (0.085)	0.038 (0.088)	
School 3	0.067 (0.083)		0.085 (0.088)
School 4	0.035 (0.085)		0.036 (0.090)
School 5	0.062 (0.082)		0.084 (0.090)
School 6	-0.038 (0.084)		-0.027 (0.090)
School 7	0.052 (0.082)	-0.006 (0.089)	
School 8	0.032 (0.080)	-0.031 (0.087)	
School 9	-0.018 (0.093)		-0.019 (0.099)
School 10	0.084 (0.092)	0.037 (0.095)	
School 11	0.042 (0.088)	-0.018 (0.094)	
School 12	0.014 (0.088)		0.014 (0.092)
School 13	0.065 (0.082)		0.058 (0.083)
School 14	0.045 (0.072)		0.048 (0.073)
School 15	0.051 (0.073)	0.001 (0.081)	
School 16	0.160 (0.098)	0.097 (0.102)	
School 17	0.074 (0.086)	0.000 (0.091)	
School 18	0.031 (0.081)		0.042 (0.088)

(continued)

Appendix Table B.6 (continued)

Variable	Parameter Estimates (Standard Errors)		
	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
School 19	0.025 (0.085)		0.034 (0.090)
School 20	0.100 (0.089)	0.036 (0.096)	
School 21	0.020 (0.076)	-0.042 (0.080)	
School 22	0.036 (0.072)		0.041 (0.074)
School 23	0.019 (0.080)		0.022 (0.085)
School 24 ^a	0.054 (0.078)	-- --	
School 25	0.090 (0.093)	0.037 (0.097)	
School 26	0.049 (0.079)		0.056 (0.085)
School 27	0.002 (0.072)	-0.043 (0.081)	
School 28	0.071 (0.079)	0.009 (0.088)	
School 29	-0.006 (0.071)		-0.003 (0.072)
School 30 ^a	-- --		-- --
School 31	0.056 (0.086)	0.005 (0.090)	
School 32	0.085 (0.084)		0.089 (0.087)
School 33	0.133 (0.086)	0.075 (0.089)	
School 34 ^a	0.039 (0.086)		0.041 (0.095)
Race/ethnicity (%)			
Hispanic	-0.047 (0.042)	-0.011 (0.061)	-0.076 (0.058)
Black, non-Hispanic	-0.035 (0.035)	-0.012 (0.051)	-0.053 (0.049)
White, non-Hispanic ^a	-- --	-- --	-- --
Other	-0.033 (0.050)	0.043 (0.071)	-0.095 (0.071)

(continued)

Appendix Table B.6 (continued)

Variable	Parameter Estimates (Standard Errors)		
	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Gender (%)			
Male	-0.011 (0.021)	-0.015 (0.030)	-0.011 (0.028)
Female ^a	-- --	-- --	-- --
Average age (years)	0.009 (0.028)	0.035 (0.040)	-0.010 (0.039)
Overage for grade ^b (%)	0.026 (0.040)	-0.017 (0.058)	0.062 (0.055)
Language other than English spoken at home (%)	0.024 (0.026)	-0.006 (0.038)	0.053 (0.037)
Home language missing (%)	-0.030 (0.090)	0.026 (0.143)	-0.060 (0.118)
Mother's education level (%)			
Did not finish high school ^a	-- --	-- --	-- --
High school diploma or GED certificate	0.004 (0.035)	-0.009 (0.051)	0.013 (0.049)
Completed some postsecondary education	-0.004 (0.035)	-0.040 (0.051)	0.027 (0.049)
Don't know	-0.005 (0.038)	-0.068 (0.054)	0.057 (0.053)
Missing	0.083 (0.099)	0.084 (0.155)	0.077 (0.130)
Father's education level (%)			
Did not finish high school ^a	-- --	-- --	-- --
High school diploma or GED certificate	0.013 (0.036)	-0.009 (0.052)	0.038 (0.049)
Completed some postsecondary education	-0.042 (0.038)	0.000 (0.056)	-0.072 (0.052)
Don't know	0.037 (0.035)	0.073 (0.050)	0.000 (0.048)
Missing	-0.052 (0.082)	-0.141 (0.120)	0.013 (0.113)
GRADE reading comprehension			
Average standard score	-0.002 (0.002)	0.000 (0.003)	-0.005 (0.003)
Sample size	2,413	1,140	1,273
Degrees of freedom	50	33	33
Mean of the dependent variable	0.584	0.602	0.567
R-square	0.012	0.014	0.017
F-statistic	0.573	0.491	0.668
P-value of F-statistic	0.993	0.993	0.925

(continued)

Appendix Table B.6 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

^aCovariates marked by '--' were not included in the regression. The site with the highest response rate was not included.

^bA student is defined as overage for grade if he or she turned 15 before the start of the ninth grade.

Appendix C

Statistical Power and Minimum Detectable Effect Size

This appendix reviews the statistical-power analysis that was conducted during the design phase of the study to determine an acceptable level of precision when estimating the impact of the ERO programs. Specifically, it reviews how the sample configuration, use of regression covariates, and other analytic assumptions would affect the precision of the impact estimates. The discussion focuses on achievement test score outcomes because of their prominence in the study.

The discussion that follows reports precision as “minimum detectable effect sizes” (MDES). Intuitively, a minimum detectable effect is the smallest program impact that could be estimated with confidence, given random sampling and estimation error.¹ This metric, which is used widely for measuring the impacts of educational programs, is defined in terms of the underlying population standard deviation of student achievement. For example, an MDES of 0.20 indicates that an impact estimator can reliably detect a program-induced increase in student achievement that is equal to or greater than 0.20 standard deviation of the existing student distribution. This is equivalent to approximately four Normal Curve Equivalent (NCE) points on a nationally norm-referenced achievement test and translates roughly into the difference between the 25th and the 31st percentile.

Unfortunately, there is no definitive standard for a policy-relevant or cost-effective MDES. A meta-analysis of treatment effectiveness studies sheds some light on this issue.² This study found that, out of 102 studies, most of which were from education research, the bottom third of the distribution of impacts ranged from about 0 to 0.32 effect size; the middle third of impacts ranged from 0.33 to 0.50; and the top third of impacts ranged from 0.56 to 1.26. Under these “rules of thumb,” an MDES of 0.32 would be considered small. More recent work by Bloom et al. suggests that a 0.32 MDES would be considered quite large when placed in the context of the growth in test scores expected over the course of a full year of schooling. Based on data from many of the most widely used standardized reading tests, they find that the expected growth in reading for ninth-grade students ranges from a 0.11 effect size to a 0.26 effect size for a full year of school.³ Documentation for the GRADE assessment that is being used for the ERO study indicates that the expected growth for ninth-grade students is equivalent to approximately a 0.07 effect size.

The ERO impact study was designed to allow an MDES of approximately 0.06 for the full sample of schools in the study and an MDES of approximately 0.10 for the groups of schools using each of the ERO program models. The estimates of minimum detectable effect sizes for the ERO study design accounted for both within-site and across-site variation in the

¹A minimum detectable effect is defined as the smallest true program impact that would have an 80 percent chance of being detected (have 80 percent power) using a two-tail hypothesis test at the 5 percent level of statistical significance.

²Lipsey (1990).

³Bloom, Hill, Black, and Lipsey (2006).

outcome in question. They also accounted for random differences between the program and control groups by including pre-random assignment reading test scores. Finally, the minimum detectable effect sizes presented in the study design were assumed to be fixed-effect estimates; that is, they did not account for variation across sites in the true impact of the program.⁴ This final assumption was justified by the fact that sites for the study were to be selected purposefully. Statistically, therefore, the results reflect the impact for the particular sample of schools in the study and should not be generalized to a broader population of similar schools.

Appendix Table C.1 shows the sample sizes resulting from various configurations of schools and student subgroups. The upper panel shows sample sizes in the ideal case that follow-up data would be available for all students in the sample. The lower panel shows sample sizes in cases where those follow-up data would be available for 80 percent of the students in the sample. Each row in the exhibit shows the sample sizes for various groupings of schools. Each column in the table shows sample sizes for potential subgroups of the targeted number of students that the study aimed to include.

There are 34 schools in the ERO study sample. Initially, the study aimed to identify approximately 110 students for each of two cohorts of ninth-graders who would be eligible and

⁴Minimum detectable effect sizes were estimated as follows:

$$MDES = 2.8 * \sqrt{\frac{\sigma_y^2(1 - R^2)}{P(1 - P)(n)(J)(\sigma_y^2 + \tau_y^2)} + \frac{\omega^2}{J(\sigma_y^2 + \tau_y^2)}}$$

where:

σ_y^2 = the (within site) variance of the outcome in question (assumed to be 1; however, by definition of effect size metric, does not affect the minimum detectable effect size);

R^2 = the explanatory power of the impact regression adjusted for pre-random assignment characteristics, i.e., the proportion of the variance in y explained by the experiment and any pre-random assignment characteristics. In order to determine an appropriate r-square, MDRC regressed ninth-grade SAT-9 achievement on eighth-grade scores for high school students in the Houston school district in 2002. The regression produced an r-square value of 0.69, which we used in our effect size calculations.

P = the proportion of students randomly assigned to the treatment group (assumed to be 0.55 based on the random assignment design for this study);

n = the number of students in each site (as listed in Appendix Table C.1);

J = the number of sites in the study (as listed in Appendix Table C.1);

τ_y^2 = the cross site variance in the mean value of the outcome measure y and calculated as 0.08 (based on an assumption that the intra-class correlation $\frac{\tau^2}{\tau^2 + \sigma^2} = 0.07$, an assumption based on MDRC's analysis of achievement data across all comprehensive non-exclusive high schools in the Houston school district);

ω^2 = the cross site variance in the true impact of the program. The minimum detectable effects sizes presented here are calculated as fixed effects estimates; that is, they do not account for cross-site variation in the true impact of the program. Thus, ω^2 is assumed to be zero.

appropriate for the ERO program. Of these, 60 students would be randomly assigned to enroll in the ERO classes, and the remaining 50 students would constitute the control group. Under these assumptions, the target sample for the first cohort of students in the ERO study was a total of 3,740 students. As discussed in Chapter 2, the actual sample for the first cohort was 2,916 students. This is closer to the sample displayed in the second column of numbers in Appendix Table C.1, which is highlighted to reflect the fact that most of the discussion will focus on the MDES estimates for this sample.

The two remaining columns in Appendix Table C.1 show sample sizes for subgroups comprising 50 percent of the target sample and 25 percent of the sample. The 25 percent subgroup (935 students), for example, is somewhat smaller than the actual number of students in the first cohort with baseline test scores that were between the fourth- and fifth-grade level (1,072 students.)

The second row of numbers in Appendix Table C.1 shows sample sizes for a subgroup of 17 schools reflecting the groups using each of the two supplemental literacy programs. It shows that the target sample for each ERO program was 1,870 students. In fact, the first cohort includes 1,385 students from the 17 schools using Reading Apprenticeship Academic Literacy and 1,531 students from the 17 schools using Xtreme Reading. These samples are closer to those shown in the second column of numbers in Appendix Table C.1. The third and fourth rows show the sample sizes for smaller subgroups of schools — for example, if the schools within each of the programs were split into two groups (approximately eight schools each) or if there were to be district-level analyses (seven of the 10 participating districts had four schools each).

The bottom panel of Appendix Table C.1 shows sample sizes that would result from follow-up data collection from 80 percent of the students in the original sample. As discussed in Chapter 2, approximately 83 percent of the students in the study sample completed the follow-up test and survey, for a respondent analysis sample of 2,413 students. The resulting samples sizes are closest to those shown in the second column of numbers in Appendix Table C.1.

Appendix Table C.2 shows how minimum detectable effect sizes for average reading achievement scores would vary among sample sizes associated with various configurations of sites and student subgroups. Again, as noted above, the highlighted column for 75 percent of the target sample closely approximates the minimum detectable effect sizes for the first cohort of students in the study sample. We now turn to the study's key impact questions.

What is the impact of supplemental literacy interventions of the type that were selected on students' reading achievement? Analyses that address this question will rely on the full sample of students across all 34 participating high schools. The second column of numbers in the bottom panel of Appendix Table C.2 indicates that the MDES for this sample would be 0.06

standard deviation if the follow-up data collection effort achieved at least an 80 percent response rate.

What is the impact of each supplemental literacy intervention on students' reading achievement? Analyses that address this question will rely on the sample of students from 17 of the 34 participating high schools. The second column of the bottom panel of Appendix Table C.2 indicates that the MDES for this sample would be 0.09 standard deviation if the follow-up data collection effort achieved at least an 80 percent response rate.

What is the impact of each supplemental literacy intervention on reading achievement for important subgroups of students or sites? In addition to questions regarding effects for the full sample of students and for students in high schools implementing each literacy intervention, the evaluation was designed to allow for the estimation of impacts for subgroups of students defined by pre-random assignment characteristics, including baseline reading test scores, whether students had been retained in a prior grade, and English language-learning status.

The last column in Appendix Table C.2 presents the estimated minimum detectable effect sizes for subgroups of students that would comprise at least 25 percent of the intended sample and approximately one-third of the actual sample. For example, students with especially low baseline test scores (between the fourth- and fifth-grade level) comprise approximately a third of the actual sample. The MDES for a subgroup of this size (approximately 935 students) would be 0.11 standard deviation units for analyses that include all 34 high schools and 0.16 for analyses that focus only on the 17 schools using one or the other of the two supplemental literacy programs.

Appendix Table C.1

**Sample Sizes, by Site and Student Subgroup Configuration,
for Full Sample and 80 Percent Subsample**

100 Percent Response Rate Sample Size				
Number of Schools	Target Sample	75 Percent of Target Sample	50 Percent of Target Sample	25 Percent of Target Sample
34	3,740	2,805	1,870	935
17	1,870	1,403	935	468
8	880	660	440	220
4	440	330	220	110
80 Percent Response Rate Sample Size				
Number of Schools	Target Sample	75 Percent of Target Sample	50 Percent of Target Sample	25 Percent of Target Sample
34	2,992	2,244	1,496	748
17	1,496	1,122	748	374
8	704	528	352	176
4	352	264	176	88

Appendix Table C.2

**Minimum Detectable Effect Sizes, by Site and Student Subgroup Configuration,
for Full Sample and 80 Percent Subsample**

100 Percent Response Rate Minimum Detectable Effect Size				
Number of Schools	Target Sample	75 Percent of Target Sample	50 Percent of Target Sample	25 Percent of Target Sample
34	0.05	0.06	0.07	0.10
17	0.07	0.08	0.10	0.14
8	0.10	0.12	0.14	0.20
4	0.14	0.17	0.20	0.29
80 Percent Response Rate Minimum Detectable Effect Size				
Number of Schools	Target Sample	75 Percent of Target Sample	50 Percent of Target Sample	25 Percent of Target Sample
34	0.06	0.06	0.08	0.11
17	0.08	0.09	0.11	0.16
8	0.11	0.13	0.16	0.23

Appendix D

ERO Implementation Fidelity

This appendix describes the development of measures based on the classroom observation data collected during site visits to the ERO high schools. The analysis of ERO program implementation fidelity in the first year of the study is based on field research visits to each of the 34 high schools during the second semester of the 2005-2006 school year. The primary data collection instrument for the site visits was a set of protocols for classroom observations and interviews with the ERO teachers. The observation protocols provided a structured process for trained classroom observers to rate characteristics of the ERO classroom learning environments and the ERO teachers' instructional strategies. All of these characteristics (referred to as "constructs") were selected for assessment because they were aligned with program elements specified by the developers and, by design, were aligned with supplemental literacy program elements that are believed to characterize high-quality interventions for struggling adolescent readers.¹ The instrument included ratings for six general instructional constructs that are common to both Reading Apprenticeship Academic Literacy and Xtreme Reading and ratings for seven program-specific constructs for each of the two interventions. The program-specific constructs reflect the distinctive components of the two ERO programs and are designated with program-specific terminology. (The observation protocols are included at the end of this appendix.)

Before conducting the classroom observation visits, observers — who were research employees of the American Institutes for Research (AIR) and MDRC who had worked previously on at least one project involving site visits — attended a two-day training to learn about the program designs and their intended implementation strategies and to learn and practice how to use the protocols. The classroom observations were conducted by two researchers (one a senior staff member with at least a master's degree, and the other a junior staff person who had at least a bachelor's degree) and captured between 160 and 180 minutes of instruction in each of the 34 high schools. The amount of observation time in each school ranged from at least two ERO classes (in schools with 80- to 90-minute class periods) and up to four ERO classes (in schools with 45-minute class periods).

Site visits were scheduled with the intent of observing classrooms across schools after similar amounts of instructional time had passed. On average, the observations occurred 21 weeks after the ERO classes started. Given that the programs ran for an average of 30 weeks, the observations occurred when the teachers had had time to cover much of the curriculum but had not yet experienced teaching all of it. The fact that the measurement of implementation fidelity is based on a single set of classroom observations also means that the measures do not capture the full range of experiences that teachers had with the programs or changes in implementation fidelity over time.

¹Biancarosa and Snow (2004).

During the visit to a given school, observers took detailed field notes, focusing on teachers' presentation of curriculum components, the flow of instruction, students' behavior and engagement, and teacher-student interactions. Each of the two observers then gave a preliminary summative rating, across all the observed classes in the school, for each of six common program constructs (used in the observations for both programs) and for each of the seven program-specific constructs (with different constructs used in observations of Reading Apprenticeship and Xtreme Reading). If the two observers gave different ratings initially, they discussed the rationale for their ratings and reached agreement about what the final ratings should be for each construct. The final rating for each construct was accompanied by a justification statement tying the observed behaviors and activities to the descriptions of the expected behaviors and activities that were used to guide the observations. The ratings from all the site visits were reviewed centrally by at least two senior members of the study team, who checked that the justifications for the ratings were grounded in the types of evidence called for in the observation protocols.

The observers used a three-category rating format for each of the general and program-specific constructs.² Although each construct was rated using criteria that were specific to that construct, the following provides a general description of the principles that were embedded in each of the three rating categories.

- **Category 3.** For each construct, classes that fell into this category included teacher behaviors and classroom activities that were well developed and highly consistent in their alignment with the intended behaviors and activities specified by the developers and described in the protocol. In these classes, teachers demonstrated confidence in what they were teaching, conveyed a thorough understanding of what was being taught conceptually and procedurally, were familiar with any materials needed, and were able to interact proactively with students who asked questions or experienced difficulty. Students appeared to be engaged in the instruction and demonstrated learning behaviors that went beyond rote performance. Teachers who fell into this category took advantage of opportunities to connect instruction to a spontaneous event or interaction in class ("a teachable moment"). If students worked independently during some of the class, they were engaged and seemed to understand the purpose of and procedures for their activity.

²In some cases, a rating of "not applicable" was used to show that the construct was not observed at all during the site visit. Two situations may have necessitated the need for this rating. First, the lesson being taught on the day of the observation did not call for attention to the construct. Second, opportunities to address a particular construct did not arise during the course of the class. Constructs with a "not applicable" rating were treated as missing data and were not given a numeric value.

- **Category 2.** For each construct, classes that fell into this category included observed teacher behaviors and classroom activities that were at least moderately aligned with the behaviors and activities specified by the developers and described in the protocols. Teachers demonstrated more than a basic understanding of what they were teaching but might not have taken full advantage of opportunities to use program materials, capitalize on “teachable moments,” or explain fully a strategy or concept. In these classes, students, while generally attending to the instruction or task at hand, did not appear intellectually engaged, and some may have been inattentive or confused.
- **Category 1.** For each construct, classes that fell into this category were not aligned with the behaviors and activities specified by the developers and described in the protocols. Teachers may have neglected opportunities to teach, may have paid only limited attention to an aspect of the program, and may not have been responsive to students’ confusion or questions. In these classes, students were sporadically engaged in the lesson, and some students may have been acting in a disruptive fashion.

There are five ways in which the study team sought reliable ratings across site visits. First, all observers were trained together to promote a common understanding of the observation process. Second, researchers went into the field in pairs with the expectation that they would collaboratively rate the implementation constructs they observed. That is, if the two observers rated a construct differently, they discussed the rating until they reached agreement about what it should be. Third, although observer pairs observed all of the participating high schools in a school district, the pair of individuals within each rating team varied across districts, thus limiting the potential for the development of particularistic understandings by a given pair of observers of how to rate the constructs. Fourth, the summative ratings from all the site visits were reviewed centrally by senior members of the study team, who checked that the justifications for the ratings were grounded in the types of evidence called for in the observation protocols. If the reviewers questioned a rating, the observers and reviewers reached a decision on keeping or changing the rating based on review of the observation data. Last, all of the site observers met as a group during the site visits to discuss the rating process and reinforce a common understanding of the relationship between the rating scale and the constructs.

Measuring the Classroom Learning Environment

As discussed in Chapter 3, the measurement of implementation fidelity focused on two key dimensions of implementation: learning environment and comprehension instruction. Ratings for the constructs were combined to calculate composite measures for each of these two

key dimensions. This section of the appendix describes how the composite measure of the learning environment dimension was calculated.

Learning Environment Composite (2 items, Cronbach's alpha = .84)

This measure was designed to measure the extent to which ERO classrooms represented learning environments believed to be conducive to the effective delivery of the core instructional strategies by the teacher and the facilitation of student and teacher interactions around the reading skills that were being taught and practiced. It was created by averaging a general instructional component measured at all 34 ERO high schools and a program-specific component measured at each set of 17 schools implementing each program.

General Instructional Learning Environment Component (2 items, Cronbach's alpha = .77)

This component is the average of two observed constructs that are part of the general instructional scales: classroom climate and on-task participation.³

Program-Specific Learning Environment Components

Reading Apprenticeship (1 item, Cronbach's alpha = na)

The program-specific component of the learning environment composite for Reading Apprenticeship schools is a single construct: social reading community. Thus the calculation of a Cronbach's Alpha is not applicable.

Xtreme Reading (2 items, Cronbach's alpha = .85)

The program-specific component of the learning environment composite for Xtreme Reading schools is the average of two constructs: classroom management and motivation and engagement.

³In the observation protocols, "motivation and student engagement" is used to describe both a general instructional construct and an Xtreme Reading-specific construct. In this discussion and the discussion in Chapter 3, the general instructional construct has been renamed "on-task participation" to distinguish it more clearly from the program-specific construct, still referred to as "motivation and student engagement."

Equations D-1 and D-2 (below) show how the constructs and components were combined to calculate the learning environment composite measures for Reading Apprenticeship and Xtreme Reading schools.⁴

$$LE_{RA} = \frac{1}{2} (\frac{1}{2} (GIC_1 + GIC_2) + (PSC_{RA1})) \quad (D-1)$$

Where:

- LE_{RA} = learning environment composite measure in a Reading Apprenticeship school
- GIC_1 = classroom climate (general instructional construct)
- GIC_2 = on-task participation (general instructional construct)
- PSC_{RA1} = social reading community (Reading Apprenticeship construct)

$$LE_{XR} = \frac{1}{2} (\frac{1}{2} (GIC_1 + GIC_2) + \frac{1}{2} (PSC_{XR1} + PSC_{XR2})) \quad (D-2)$$

Where:

- LE_{XR} = learning environment composite measure in an Xtreme Reading school
- GIC_1 = classroom climate (general instructional construct)
- GIC_2 = on-task participation (general instructional construct)
- PSC_{XR1} = classroom management (Xtreme Reading construct)
- PSC_{XR2} = motivation and engagement (Xtreme Reading construct)

Measuring Reading Comprehension Instruction

This section of the appendix describes how the composite measure of the second key implementation dimension, comprehension instruction, was calculated.

Comprehension Instruction Composite (2 items, Cronbach's alpha = .72)

This measure was designed to measure the quality of the reading comprehension instruction in each ERO school. As with the learning environment composite measure, it was created by averaging a general instructional component measured at each of the 34 ERO high

⁴In these equations, "LE" stands for learning environment; "RA" and "XR" stand for Reading Apprenticeship and Xtreme Reading respectively; and "GIC" and "PSC" stand for general instructional construct and program-specific construct respectively.

schools and a program-specific component measured at each school — the Reading Apprenticeship component at each of the 17 Reading Apprenticeship schools and the Xtreme Reading component at each of the 17 Xtreme Reading schools.

General Instructional Comprehension Instruction Component (2 items, Cronbach's alpha = .81)

This component is the average of two observed constructs that are part of the general instructional scales: comprehension and metacognition.

Program-Specific Comprehension Instruction Components

Reading Apprenticeship (5 items, Cronbach's alpha = .70)

The program-specific component of the comprehension instruction composite for Reading Apprenticeship schools is the average of five constructs observed at and averaged for each school: metacognitive conversations, silent sustained reading, content/theme integration, writing, and integration of curriculum strands.

Xtreme Reading (2 items, Cronbach's alpha = .50)

The program-specific component of the comprehension instruction composite for Xtreme Reading schools is the average of two constructs: curriculum-driven (or systematic) instruction and needs-driven (or responsive) instruction. The curriculum-driven instruction construct is the average of three subconstructs: structured content, research-based methodology, and connected scaffolded and informed instruction (Cronbach's alpha = .74). The needs-driven instruction construct is the average of two subconstructs: student accommodations and feedback to students (Cronbach's alpha = .71).

Equations D-3 and D-4 (below) show how the constructs and components were combined to calculate the comprehension instruction composite measures for Reading Apprenticeship and Xtreme Reading schools.⁵

⁵In these equations, "CI" stands for comprehension instruction; "RA" and "XR" stand for Reading Apprenticeship and Xtreme Reading respectively; and "GIC" and "PSC" stand for general instructional construct and program-specific construct respectively.

$$CI_{RA} = \frac{1}{2} (\frac{1}{2} (GIC_1 + GIC_2) + \frac{1}{5} (PSC_{RA1} + PSC_{RA2} + PSC_{RA3} + PSC_{RA4} + PSC_{RA5})) \quad (D-3)$$

Where:

- CI_{RA} = comprehension instruction composite measure in a Reading Apprenticeship school
- GIC_1 = comprehension (general instructional construct)
- GIC_2 = metacognition (general instructional construct)
- PSC_{RA1} = metacognitive conversations (Reading Apprenticeship construct)
- PSC_{RA2} = silent sustained reading (Reading Apprenticeship construct)
- PSC_{RA3} = content/theme integration (Reading Apprenticeship construct)
- PSC_{RA4} = writing (Reading Apprenticeship construct)
- PSC_{RA5} = integration of curriculum strands (Reading Apprenticeship construct)

$$CI_{XR} = \frac{1}{2} (\frac{1}{2} (GIC_1 + GIC_2) + \frac{1}{2} (PSC_{XR1} + PSC_{XR2})) \quad (D-4)$$

Where:

- CI_{XR} = comprehension instruction composite measure in an Xtreme Reading school
- GIC_1 = comprehension (general instructional construct)
- GIC_2 = metacognition (general instructional construct)
- PSC_{XR1} = systematic instruction (Xtreme Reading construct; the average of measures of structured content, research-based methodology, and connected, scaffolded, informed instruction)
- PSC_{XR2} = responsive instruction (Xtreme Reading construct; the average of measures of student accommodations and feedback to students)

Categorizing Implementation Fidelity

This section of the appendix discusses briefly how schools were categorized based on the ratings calculated for each of the 34 participating high schools on the implementation fidelity of their classroom learning environment and for the implementation fidelity of their comprehension instruction. Each overall rating ranged between 1 and 3, and was rounded to the nearest tenth of a point. Based on the composite ratings for each of the two program dimensions — learning environment and comprehension instruction — the implementation fidelity for each dimension was classified as “well aligned,” “moderately aligned,” or “poorly aligned” to the models specified by the program developers.

A dimension rated at the level of “well-aligned” implementation fidelity received an average composite rating of 2.0 or higher. A dimension rated at the level of “moderately

aligned” implementation fidelity received an average composite rating between 1.5 and 1.9. A dimension rated at the level of “poorly aligned” implementation fidelity received an average composite rating that fell below 1.5 (on a scale ranging from 1 to 3).

The top two panels of Appendix Table D.1 provide a summary of the number of schools whose composite rating on the classroom learning environment and comprehension instruction dimensions fell into the well-aligned, moderately aligned, and poorly aligned categories of fidelity. These panels are the same as the top two panels of Table 3.5 in Chapter 3. The bottom panel of the table clusters schools based on their level of implementation fidelity across both dimensions. This panel clusters the schools into more categories of combined implementation fidelity than the same panel in Table 3.5.

The Enhanced Reading Opportunities Study

Appendix Table D.1

**Number of ERO Classrooms Well, Moderately, or Poorly Aligned
to Program Models on Each Implementation Dimension,
by ERO Program**

Implementation Dimension	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
<u>Learning environment</u>			
Well-aligned implementation (composite rating is 2.0 or higher)	26	14	12
Moderately aligned implementation (composite rating is 1.5-1.9)	4	2	2
Poorly aligned implementation (composite rating is less than 1.5)	4	1	3
<u>Comprehension instruction</u>			
Well-aligned implementation (composite rating is 2.0 or higher)	16	7	9
Moderately aligned implementation (composite rating is 1.5-1.9)	9	4	5
Poorly aligned implementation (composite rating is less than 1.5)	9	6	3
<u>Combined dimensions</u>			
Well-aligned implementation on both dimensions	16	7	9
Well-aligned implementation on learning environment only ^a	10	7	3
Well-aligned implementation on comprehension instruction only	0	0	0
Moderately aligned implementation on both dimensions	2	1	1
Poorly aligned implementation on learning environment only	1	0	1
Poorly aligned implementation on comprehension instruction only ^a	6	5	1
Poorly aligned implementation on both dimensions	3	1	2
Sample size	34	17	17

(continued)

Appendix Table D.1 (continued)

SOURCES: MDRC and AIR calculations from classroom observation data.

NOTES: Implementation with a composite score of less than 1.5 for a given dimension was deemed to be at the beginning stages of development. The implementation for these dimensions was designated as poorly aligned with the program models.

Implementation with composite scores between 1.5 and 1.9 for a given dimension exhibited at least moderate development in some areas while being at the beginning stages of development in other areas. The implementation for these dimensions was designated as moderately aligned.

Implementation with scores of 2.0 or higher for a given dimension exhibited well-developed fidelity on several areas and at least moderate development in most other areas. The implementation for these dimensions was designated as well aligned.

^aFour Reading Apprenticeship schools were designated as being well aligned in terms of learning environment and poorly aligned in terms of comprehension instruction. Thus, these schools are counted in two rows in the bottom panel of the table.

Observation Protocols

Table of Contents

General Instruction Scales	154
Reading Apprenticeship Fidelity Scales.....	159
Xtreme Reading Fidelity Scales	167

Enhanced Reading Opportunities Program

General Instruction Scales

Area of interest	Basic Literacy Skills (Advanced phonics and decoding, fluency)
Description	
<p>0. Not applicable. During the observed class period(s), students do not demonstrate a need for instruction in basic literacy skills.*</p> <p>1. During the observed class period(s), instruction does not reflect teacher recognition of a <i>demonstrated</i> student need for increased understanding of basic literacy skills. The teacher may not recognize or acknowledge this need for practice of basic literacy skills OR these skills are addressed but in a very cursory manner (e.g., students are told to “sound out” words they don’t know).</p> <p>2. During the observed class period(s), instruction reflects teacher recognition of student difficulty with basic literacy skills; however, instruction is not really well developed. For example, fluency and decoding skills may be practiced in a “skill and drill” manner and never applied to authentic texts. As other examples, instruction may not be differentiated to meet individual student needs, OR the teacher may provide insufficient practice opportunities.</p> <p>3. During the observed class period(s), instruction reflects teacher recognition of student difficulty with basic literacy skills <i>and</i> the instruction is provided in a manner that meets student needs. Such instruction could take several forms. For example, instruction could be differentiated for individual students, OR ample practice opportunities could be provided for those who need it, in order to facilitate increased decoding and fluency abilities, as well as the ability to apply these skills to make meaning of text. This could be evidenced by students learning or applying a systematic approach for decoding unknown words as they read a piece of literature).</p>	

*A demonstrated need could be manifested in the form of student difficulties with decoding words, or students reading haltingly or without expression.

Area of interest	Vocabulary
Description	
<ol style="list-style-type: none"> <li data-bbox="235 367 1424 441">1. There was no opportunity for vocabulary instruction to occur during the observed class period(s). OR Students are engaged in a few vocabulary development activities, but these activities are largely superficial in nature. Vocabulary is not connected to student texts or writing. Such instruction could take the form of rote vocabulary learning methods, OR vocabulary instruction that occurs out of textual context. For example, students may be asked to look up the definitions of words in the dictionary to discover meanings. <li data-bbox="235 630 1424 808">2. Students are engaged in some vocabulary activities, but these activities are not fully developed. For example, the teacher may employ definitional and contextual information for presenting words but gives little attention to linking words to prior experiences OR to teaching strategies to help students figure out the meaning of words on their own (e.g. identifying root word, using context clues, etc). <li data-bbox="235 819 1424 1018">3. Students are engaged in vocabulary instruction that is integrated throughout instruction, and multiple vocabulary strategies are used. Instruction provides students with strategies that help them to independently derive the meaning of unfamiliar words. For example, instruction may focus on using strategies to identify new words and building context for new words and concepts. Repetition and both direct and indirect techniques for teaching vocabulary may be utilized. 	

Area of interest	Comprehension
Description	
<p>1. There was no opportunity for comprehension instruction to occur during the observed class period(s). OR</p> <p>Few opportunities are provided for students to obtain meaning from text, and comprehension strategies are addressed in a basic or superficial manner. For example, the teacher or the students may expend little effort to understand the substance of what is being read. Instruction may not be focused on reading text and meaning-making, or the teacher may do very little modeling and direct instruction of comprehension strategies. The teacher may make little or no efforts to monitor student comprehension of text.</p> <p>2. Some opportunities are provided for students to try to obtain meaning from text, but comprehension strategies are not fully developed. For example, students may make some attempts to make sense of difficult or unfamiliar text, but they give up easily when they don't understand. As another example, the teacher may make some attempts to model critical thinking strategies, but direct instruction is limited to teaching basic comprehension strategies (e.g., making predictions, identifying main characters and setting, and summarizing, distinguishing between fact and opinion). The teacher may monitor or probe for student comprehension but does not necessarily use this information to target or enhance specific comprehension skills during the class period.</p> <p>3. There are substantial opportunities and various approaches for students to try to obtain and validate meaning from text. Most students, for most of the time, are trying to derive meaning from the texts that they read and have concrete strategies for doing so. Opportunities for the development of student reading skills could be evidenced by teacher use of modeling and direct instruction to teach strategies and thought processes, and emphasis of critical thinking. The teacher may also encourage or facilitate purposeful student discussion and interaction with text. For example, the teacher may activate students' prior knowledge and encourage higher-order thinking. Instructional content may include components of text structure, both generically and with specific reference to content-area learning. Another example of substantial comprehension instruction could include teacher monitoring or probing for student comprehension, followed by teaching or reflecting on strategies to enhance student comprehension abilities.</p>	

Area of interest	Metacognition
Description	
(Note: In a successful class, this becomes less visible towards the end of the year as students internalize these procedures.)	
<ol style="list-style-type: none"> <li data-bbox="233 401 1406 632">1. Little metacognitive work is apparent, and overall, metacognitive skills are not being <i>developed</i> through instruction or conscious practice. In some cases, students may be taught strategies to monitor their own reading, recognize faulty comprehension, and apply “fix-up” strategies; but these strategies are not explored. For example, the teacher either does not address metacognitive strategies (e.g., self-monitoring of reading may not be taught at all) or does so in a very limited or superficial, contrived manner (e.g., teacher and students are most often “going through the motions”). <li data-bbox="233 653 1419 848">2. Instruction incorporates some development of metacognitive strategies and opportunities for student practice of metacognition, either through spoken or written expression, but these may not be fully developed. For example, instruction could include the use of “think alouds” to model strategies, self-correct, and make connections to prior knowledge. While some of the metacognitive activities flow naturally, others may appear to be forced (teacher or students appear to be “going through the motions”). <li data-bbox="233 869 1414 1064">3. Use of metacognitive strategies is pervasive and integrated throughout instruction. Instruction includes teacher modeling of strategies and multiple opportunities for student practice of thinking aloud through spoken or written expression with multiple forms of text. Throughout the majority of metacognitive activities, the teacher monitors and guides students in their thought processes. In addition, the majority of the metacognitive activities are conducted in a natural and thoughtful manner. 	

Area of interest	Classroom Climate and Social Support for Learning
Description	
<ol style="list-style-type: none"> <li data-bbox="233 338 1398 533">1. The classroom environment seems disrespectful and chaotic. Students interrupt each other and interfere with one another's efforts to learn. For example, students may engage in or experience taunts, occasional threats, or slurs about themselves or backgrounds. The teacher does little, if anything, to counteract these problems. Students have little opportunity to work together (either in pairs or small groups) towards a common goal; limited student voluntary participation is observed. <li data-bbox="233 558 1390 753">2. The classroom environment seems somewhat respectful, but there are some instances of disruptive or disrespectful student behavior. For example, the teacher may attempt to provide a safe environment and/or provide some instruction on how to work together, but students occasionally engage in and/or experience put-downs, taunts, even occasional threats or slurs about themselves or backgrounds. The teacher rectifies the problem on a situation-by-situation basis. The teacher may or may not encourage reluctant students to participate in discussions. <li data-bbox="233 779 1390 905">3. The classroom environment appears to reflect mutual and widespread respect between teachers and students. The classroom is characterized by few, if any, taunts and primarily polite, appropriate interactions among students and between students and teacher. For the majority of instruction, both teacher and students solicit and welcome contributions from all students. 	

Area of interest	Motivation and Student Engagement
Description	
<ol style="list-style-type: none"> <li data-bbox="233 1148 1414 1274">1. Disruptive or passive disengagement; most students are frequently off-task, as evidenced by either gross inattention or serious disruptions. For substantial portions of time, many students are either off-task or nominally on-task but not trying very hard. Students could appear to be lethargic and disinterested in class activities or they might be actively misbehaving. <li data-bbox="233 1299 1382 1390">2. Sporadic or episodic engagement; most students, some of the time, are engaged in class activities. Engagement may be uneven, mildly enthusiastic or dependent on frequent prodding from the teacher. <li data-bbox="233 1415 1409 1476">3. Engagement is widespread; most students are on-task most of the time pursuing the substance of the lesson. The majority of students seem to be taking the work seriously and trying hard. 	

Enhanced Reading Opportunities Program

Reading Apprenticeship Academic Literacy Fidelity Scales

Core Principle # 1

Social Reading Community

A *Social Reading Community* is established so that students can work collaboratively with their teacher and peers to derive meaning and pleasure from text.

- A safe and nurturing classroom environment is established.
- Well-established classroom routines foster peer interaction.
- Through teacher modeling, students are encouraged to recognize and use the diverse perspectives and resources brought by each member of the class.
- Students are encouraged to share their confusion and difficulties with texts, without fear of embarrassment or punishment.
- Teacher actively listens to and responds to students' comments in teacher-facilitated conversations; over the course of the year, students increasingly contribute to and guide whole-class conversations and activities.
- Teacher takes steps to encourage active student participation and to invite diverse responses.
- Teacher shares his or her own struggles, satisfactions and reading processes.

Fidelity Scale

1. The classroom environment does not promote an open exchange of student ideas about text. The teacher may do little or no modeling of such interaction.

Such an environment could be characterized by little or no student sharing related to the evaluation or generation of meaning from text. Many students may appear to be reluctant to participate in discussions related to text most of the time. The teacher may have to work extremely hard to get students to interact about text meaning, or prompting by the teacher to encourage student conversations about literature is ineffective.

Instruction in this category could also be characterized by students ridiculing their peers when they acknowledge confusion about text. The teacher may ignore student attempts to express confusion or may not model respect for the varied perspectives and ideas of all members of the classroom community.

2. In general, the classroom environment appears to be a safe place to interact and share ideas about text. The teacher occasionally models appropriate ways for sharing ideas about text.

A moderately developed social reading community could be characterized by discussions about text that are primarily teacher-directed during the majority of the instructional period. Classroom routines for peer interaction may not be fully developed. Some students may appear to be hesitant to volunteer their own ideas or confusion about text. As another example, the teacher may actively listen to student responses and attempt to elicit a variety of responses from all members of the reading community, but he or she has trouble engaging the majority of students in discussion of literature or of text meaning.

3. A safe and nurturing environment is established for students to share ideas about text. When necessary, the teacher models a process for sharing ideas about text.

This social reading community could be characterized by frequent student participation. The majority of students contribute to or guide whole-class or group conversations and activities related to literature and other forms of text. They may also volunteer confusion and difficulties with texts. A positive social reading community could also be evident during teacher-facilitated conversations that encourage active participation from all members of the classroom community.

Core Principle # 2	Metacognitive Conversation
---------------------------	-----------------------------------

Metacognitive Conversation is a regularly occurring routine which is evident in RAAL classroom work and interactions:

- Students are taught to use classroom inquiry to generate a repertoire of specific comprehension and problem-solving strategies.
- Through ongoing conversations rooted in text, students learn to ask critical questions about content, purpose, and perspective.
- Students are encouraged to draw on strategic skills they use in out-of-school settings to assist them in solving comprehension problems.
- Students recognize that confusion can be a starting place for collaborative problem-solving aimed at deriving meaning from difficult text.
- Students have many opportunities to practice sharing and exploring their thinking about texts *with peers*; these peer-guided metacognitive conversations become more text-based and sophisticated over the course of the academic year.
- Students monitor their own mental processes for reading and adjust as needed.*
- During discussions, teacher probes for deeper student responses to enrich student learning and thinking processes.
- Teacher models metacognitive process (e.g. Thinking Aloud, Talking to the Text) *and* follows through on such practices with continued modeling and appropriate scaffolding to ensure that streams of thought are fully developed.

Fidelity Scale

1. Students are not explicitly taught a variety of comprehension and problem-solving skills. Students are primarily engaged in instruction that is aimed at uniform understandings and single correct responses.

For example, there is little evidence that reading comprehension difficulties are seen as valuable starting points for collaborative problem-solving. Students have few opportunities to practice discussing their thought processes about reading and to ask critical questions about text content. Students do not volunteer to discuss confusion about text. Students are never or rarely asked to make connections to strategic skills they use in out-of-school settings to assist them in solving comprehension problems.

As another example, the teacher does not model metacognitive strategies, or does not provide scaffolds for students to practice and apply such strategies. Instruction that falls into this category could be characterized by teacher attempts to model the use of metacognitive strategies that are largely unsuccessful or ineffective.

2. Students are taught comprehension and problem-solving skills, and at least one major classroom activity provides students with an opportunity to discuss their cognitive processes.

For example, some but not all students may share reading difficulties and confusions and collaborate in problem solving. Instruction could include opportunities for students to share problem solving and strategic skills from their lives outside of school.

Instruction could also include teacher or student engagement in discussion or assessment of the effects of particular reading processes. While the teacher occasionally models metacognitive strategies or probes for deeper student responses in relation to text, only minimal attempts are made to follow through with additional modeling or appropriate scaffolds to ensure that thought streams are fully developed and transparent.

* While we are including this bullet in the general description of the principles, we will not include in the fidelity scales as this is a “high inference” item and is not easily observable.

3. Students are taught a variety of comprehension and problem-solving skills, and they actively contribute to or guide metacognitive conversations. Such conversations are predominantly text-based.

For example, many students routinely make connections to strategic skills they use in out-of-school settings to assist them in solving comprehension problems. Students may also share their confusion with text as a basis for comprehending challenging text.

As another example, the teacher frequently and authentically models metacognitive strategies (such as using confusion as a point to generate meaning) or probes for deeper student responses in relation to text. Initial modeling is followed by additional modeling and/or appropriate scaffolds aimed at ensuring that thought streams are fully developed and transparent.

Core Principle # 3**Silent Sustained Reading**

Silent Sustained Reading is a well-established routine in which personal inquiry and peer social interaction is used to build motivation and extend students' interest to new books and genres.

- Students are encouraged to explore their own preferences and reactions to books.
- Students routinely discuss SSR books with classmates in both informal and occasionally formal activities (i.e. “book talks”).
- Students set goals for their reading development and assess their own performance in meeting those goals (in terms of amount and range of books read, persistence, and fluency).
- Students practice metacognitive routines, language study, and cognitive strategies as they read SSR books.
- Teachers routinely provide support and show interest in students’ SSR in both informal and formal activities, e.g., individual conferencing, written feedback in reading logs, sharing their own SSR books and reading processes.

Fidelity Scale

1. Either SSR did not take place during the observed class period(s).
OR

Instructional time may be allocated for SSR, but this does not seem to be a developed routine. Instruction could be characterized either by little engagement in SSR or by some engagement in SSR that is not deep or broad. SSR may be a largely individual activity. For example, teachers may not help students select books and may in fact be disengaged from the class doing unrelated activities (e.g. grading papers). As another example, there may be little collaboration on comprehension problems or sharing of reading processes. Students do not have much opportunity to practice metacognitive routines, conduct language study, or do logging, goal-setting, or sharing related to SSR books.

2. The majority of students engage in independent reading during SSR. There is some exploration of SSR reading experiences but the routine is not fully developed. Instruction could be characterized by a few instances of student discussion of reading processes and sharing related to SSR books, personal goal-setting, or writing. As another example, teacher may provide some support of SSR by assisting students in selecting books that reflect their identities as readers, or by engaging in formal or informal feedback activities such as individual conferences to discuss their SSR books and written feedback in student reading logs.
3. Students are engaged in reading SSR books and in reflecting on them either in journals or metacognitive logs or through conversations with peers. In this category, SSR routinely involves the class community in metacognitive conversation, sharing reading strategies and examples for language study. Students set increasingly challenging goals for SSR and monitor their progress. Instruction could also be characterized by demonstrated teacher interest in SSR through both formal and informal activities. For example, the teacher may hold individual conferences with students to discuss their SSR books or provide written feedback in student reading logs.

Core Principle # 4**Language Study**

Language Study is routinely integrated into varied literacy experiences in the RAAL classroom in both explicit and implicit ways:

- Language study activities engage students in and focus on finding and analyzing patterns at the word, sentence, and text levels.
- Students “nominate” challenging words, phrases, and sentences from their own SSR reading and/or from class readings for analysis by the whole class.
- Students build personal dictionaries of vocabulary words, drawing from key conceptual words taught explicitly as well as from words they encounter in their SSR reading.
- Teachers routinely take advantage of informal opportunities to support academic language development, e.g., by using interesting and playful language, gracefully reframing or elaborating student thinking using academic language. (S: You could tell that was going to happen. T: It really foreshadowed the tragic ending, didn't it?)
- In planning lessons, teachers analyze texts for potential language learning opportunities, and plan language study to take advantage of these.*

Fidelity Scale

1. Language Study did not take place during the observed class period(s).

OR

The teacher makes minimal attempts to incorporate language study into instructional activities, but these opportunities are not well developed. For example, the teacher may identify important vocabulary in class and either define or ask students to define the new words; however, little instructional attention is given to the structural features of words, phrases, or texts.

2. The teacher draws students' attention to the structure of language in various course texts at the morphological, word, phrase, sentence, and discourse levels, but instruction in language study is not deep or pervasive. For example, the teacher may incorporate aspects of language study into instruction frequently but it does not appear to be consistent (part of formal instruction and informal opportunities). As another example, there may be evidence that students keep their own word lists in notebooks, but there may be little focus on students' learning to clarify the meaning of unknown words.
3. The teacher provides instruction in the structure of language in various course texts, paying attention to morphological, word, phrase, sentence, and discourse. The teacher takes advantage of informal opportunities to support academic language development. For example, the teacher uses interesting and playful language or attempts to reframe or elaborate student thinking using academic language. As another example, students keep word lists and routinely identify key words and work to clarify word meaning as they read and work with peers. Instruction could also be characterized by student identification of language for study or student engagement in class or small group analysis of challenging words, sentences, or text passages.

* While we are including this bullet in the general description of the principles, we will not include in the fidelity scales as this is a “high inference” item and is not easily observable.

Core Principle # 5**Content and Theme**

The *Content and Theme* of each of the four thematic units* in the RAAL curriculum are integral to classroom activities and discussions:

- Students practice a variety of comprehension strategies in the context of the texts and genres presented in each of the four thematic units.
- Students are encouraged to draw on their interests in larger social, political, economic, and cultural issues as they read and discuss the texts in each thematic unit.
- Students explore personal motivations and identities as readers in relation to the four thematic units.
- Students practice analyzing and synthesizing information and ideas across multiple texts and conversations in relation to the overarching themes of the four units.
- The teacher provides instruction and support for reading the complex academic materials associated with each of the four units occurs in the classroom; reading is not merely assigned and reviewed.
- Students learn and practice academic discourse (e.g., providing evidence to support thinking, interrogating author bias) appropriate for each of the four thematic units.

Fidelity Scale

1. For the majority of the instruction period, the focus of instruction does not center on the content or theme of the current unit. If the content or theme is addressed, the class engages in only tangential discussion of the materials at hand. The teacher makes no attempt to redirect or reorient students to material relevant to current thematic unit.
2. Much of the instruction is focused on the theme of the current unit but some opportunities for integrating the overarching theme with instruction are lost. For example, students may practice a comprehension strategy in the context of the texts and genres presented in this unit, but they do not draw on their own interest in larger social or cultural issues related to the theme. As another example, students may explore personal motivations or identities related to the theme but the teacher may not provide support for reading the academic materials associated with the unit. In this category, some instruction may occur with no reference to the theme.
3. The majority of instruction focuses on text and materials relevant to the theme, and the teacher provides ample support for reading complex academic materials within the current thematic unit. For example, students have multiple or extended opportunities to practice comprehension strategies specific to the context of the texts and genres presented in this unit. As another example, students explore their personal motivations and identities in relationship to the unit and draw on their interests in larger social, political, economic, and cultural issues. Students may analyze or synthesize information across multiple texts, or they may practice academic discourse appropriate for the unit.

* The four thematic units of the RAAL curriculum consist of Unit 1: Reading Self and Society; Unit 2: Reading History; Unit 3: Reading Science; and Unit 4: Reading Media.

Core Principle # 6**Writing**

Instruction provides on-going support for *writing to learn as well as learning to write* in the RAAL classroom:

- Students are explicitly taught writing processes and the structures of particular written forms through formal writing assignments that culminate each of the four thematic units.
- Instruction and support for writing and writing processes occur in the classroom; writing is not merely assigned and graded.
- Students use writing to support their learning of thematic content through a variety of tools, including dual entry journals, graphic organizers, interactive notebooks, personal dictionaries, word and sentence analysis notes, and reflective letters.
- Students use writing as a tool for increasing their comprehension of challenging texts (e.g., students write in metacognitive logs and practice the metacognitive routine of "talking to the text" in writing).

Fidelity Scale

1. The observed class period(s) did not include a writing component.

OR

Students are not explicitly taught writing processes or about the structures of particular written forms. For example, writing assignments may be given to students, but they never receive guidance on the writing process. Instruction could alternatively be characterized by a lack of opportunities for students to use writing to support their learning of thematic content or to increase comprehension of text. Metacognitive logs may be used, but appear to be used in a very rote way (students write a simple sentence or two and these are not explored further).

2. Students engage in at least one activity where they are developing writing skills and using writing to support their learning of thematic content, but one aspect is developed in greater depth than the other. For example, instruction on learning to write may be emphasized (the writing process and the structures of particular written forms) without a lot of attention to the content of the writing. As another example, thematic content may be explored through writing tools such as dual entry journals, metacognitive logs, graphic organizers, interactive notebooks, personal dictionaries, word and sentence analysis notes, and reflective letters; but the writing process is not fully explored or developed.
3. Explicit instruction is provided in the writing processes and the structures of particular written forms related to the thematic unit; the two skill/strategies are developed hand in hand. Students use writing as a tool for increasing their comprehension of challenging texts. For example, students write in metacognitive logs and practice the metacognitive routine of "talking to the text" and hone their writing skills in the process. Students may also learn to write and use writing to support their learning of thematic content through other tools, including dual entry journals, graphic organizers, interactive notebooks, personal dictionaries, word and sentence analysis notes, and reflective letters.

Core Principle # 7**Integration of the Curriculum Strands**

The teacher integrates the five RAAL Curriculum Strands* during literacy instruction

- Students are simultaneously engaged in at least two of the strands at any given time.
 - For example, while focusing on *Metacognitive Conversation* in discussing how students solved comprehension problems reading a piece in the anthology, the teacher might integrate *Language Study* by providing a mini-lesson on roots, prefixes and suffixes in helping students clarify the meaning of an unfamiliar word.
 - For another example, the teacher might integrate Writing and Content and Theme through student discussion and writing about the “essential questions” in any of the four thematic units.

Fidelity Scale

1. The teacher does not integrate curriculum strands in any of the major instructional activities.
OR

The teacher occasionally integrates two of the curriculum strands, but does not do so in a natural manner. For example, coherent connections between course themes, language study, metacognitive conversation and strategies, independent reading experiences, and/or writing are not evident throughout the majority of instruction.

2. For at least one major activity, the teacher integrates at least two strands smoothly; instruction in each of the strands is improved upon by instruction in the other. *For example*, while focusing on *Metacognitive Conversation* in discussing how students solved comprehension problems, the teacher might integrate *Language Study* by providing a mini-lesson on roots, prefixes and suffixes in helping students clarify the meaning of an unfamiliar word. During the remainder of instruction, the teacher may refer to one or more of the curriculum strands but only in passing, or without coherently integrating them with other strands.

As another example, the teacher successfully focuses on two of the strands for the majority of the instruction but does not make attempts to integrate any remaining strands.

3. The teacher finds multiple opportunities to integrate several of the five strands “fluently” and appropriately. At least two different strands appear to be seamlessly integrated at any given time. For example, the teacher recognizes and makes use of opportunities to make natural and meaningful connections between and among course themes, language study, metacognitive conversation and strategies, independent reading experiences, and writing.

* The five strands of the RAAL Curriculum consist of Metacognitive Conversation, Silent Sustained Reading, Language Study, Content/Theme, and Writing

Enhanced Reading Opportunities Program

Xtreme Reading Fidelity Scales

Core Principle # 1	Responsive Instruction
---------------------------	-------------------------------

Instruction is *responsive* to unique student needs to “personalize teaching and learning.”

- *Assessment*: Ongoing, informal assessment is used to monitor students’ performance to determine if instructional objectives are being met and strategies are being mastered.*
- *Accommodations (1.a)*: Students begin learning reading strategies using materials at their reading level. They gradually work up through the reading levels across the school year.
- *Feedback (1.b)*: Corrective and elaborative feedback is provided to help students better understand how to improve their performance of skills and strategies. Feedback helps students recognize correct practices, as well as patterns of errors, and target improvement in specific areas. Six steps for providing feedback are recommended:
 - Teacher tells students what they have done well.
 - Teacher helps students recognize and categorize errors made during practice attempts, in order to better understand their performance.
 - Teacher re-teaches one of the error types at a time (through explaining, modeling).
 - Teacher watches student practice and provides feedback.
 - Teacher asks student to paraphrase main elements of feedback.
 - Teacher prompts student to set goals for next practice attempt.

Fidelity Scale: (Core Principle 1.a: Accommodations)

- | |
|--|
| <ol style="list-style-type: none"> 1. Accommodations were not apparent during the observed class period(s).
OR
The teacher seems unaware of or unable to determine whether instructional objectives are being met and strategies are being mastered. For example, students are provided few instructional materials that match their reading level. Materials appear to be either too challenging or too easy for the majority of the students. 2. The teacher appears to be able to provide appropriate instruction to students making expected progress but appears unaware of or unable to determine appropriate instruction for students failing to make adequate progress or for students advancing rapidly through the curriculum. For example, while some students are being instructed in materials that match their reading level, the materials appear to be either too difficult or too easy for others. 3. The teacher appears to be aware of individual student needs and is able to differentiate instruction accordingly. For example, most students have been provided with instruction and are learning reading strategies using materials at their reading level. |
|--|

* While we are including this bullet in the general description of the principles, we will not include in the fidelity scales as this is a “high inference” item and is not easily observable. Assessment is addressed in the teacher interview, and teachers will be asked to describe their use of assessments to make instructional decisions.

Fidelity Scale (Core Principle 1.b: Feedback)

1. The teacher does not provide feedback to students or does so rarely. The teacher does not appear to monitor student work and performance. In general, students are expected to practice skills and strategies independently, without teacher input.
2. While the teacher occasionally provides corrective feedback to students on their practice attempts, feedback is not elaborative or mainly highlights the negative. In general, the teacher engages in only one or two of the feedback strategies outlined in the Xtreme Reading Program (telling students what they have done well, helping students to recognize and categorize errors made during practice attempts, reteaching one of the error types at a time through modeling and explaining, watching students practice, asking students to paraphrase main elements of feedback, and prompting students to set goals for their next practice attempt). There is little follow-up with students to ensure understanding so that they may improve on their next practice attempt and obtain mastery of the skill/strategy.
3. Corrective and elaborative feedback is provided to help students better understand how to improve their performance of skills and strategies. The teacher provides feedback using most or all of the strategies outlined in the Xtreme Reading Program (telling students what they have done well, helping students to recognize and categorize errors made during practice attempts, reteaching one of the error types at a time through modeling and explaining, watching students practice, asking students to paraphrase main elements of feedback, and prompting students to set goals for their next practice attempt). The teacher follows up with students to ensure understanding so that they may improve on their next practice attempt and move toward mastery of the skill/strategy.

Core Principle # 2**Systematic Instruction**

Instruction is *systematic* in nature; that is, the information (skills, strategies, and content) taught, the sequence of instruction, and various activities and materials used are carefully planned in advance of delivering instruction. Systematic instruction is to be carefully structured, connected, and scaffolded; and it should be informative.

- *Structured Content (2.a)*: Instructional content is comprised of instruction in reading strategies (e.g., vocabulary, word-identification, self-questioning, visual imagery, paraphrasing, and inferencing) and other instructional programs that support strategy instruction (ACHIEVE Skills, SCORE Skills, Talking Together, Possible Selves). Each reading strategy is divided into smaller steps/segments.
- *Research-based instructional methodology (2.b)*: Each strategy is taught using an eight-stage methodology. On each day that a reading strategy is taught, the learning activities are associated with at least one of these stages. The stages include: Describe, Model, Verbal Practice, Guided Practice, Paired Practice, Independent Practice, Differentiated Practice, and Generalization.
- *Connected Instruction (2.c)*: Teacher purposefully shows students how new information is related to skills, strategies, or content that has been previously learned, as well as to those that will be learned in the future. Course and Unit Organizers are provided to students to introduce main ideas and to demonstrate how critical information and concepts are related.
- *Scaffolded Instruction (2.c)*: Instruction moves from teacher-mediated to student-mediated across the course of instruction in one strategy. When a new strategy is introduced, multiple instructional supports (modeling, prompts, direct explanations, targeted questions, relatively basic tasks) are initially provided by the teacher. These instructional supports are gradually reduced as the student becomes more confident and begins to move toward mastering the targeted objectives.
- *Informative Instruction (2.c)*: Teacher informs students about how the learning process works and what is expected during instruction. Teacher ensures that students understand how they are progressing, how they can control their own learning at each step of the process, and why this is important.

Fidelity Scale (Core Principle 2.a: Structured Content)

1. There is little or no evidence that that the teacher is providing instruction in any of the reading strategies outlined in the Xtreme Reading curriculum (e.g., vocabulary, word-identification, self-questioning, visual imagery, paraphrasing, and inferencing) and other instructional programs that support strategy instruction (ACHIEVE Skills, SCORE Skills, Talking Together, Possible Selves). For example, the teacher appears to be using alternative instructional materials (materials outside of the Xtreme Reading curriculum).
2. While the teacher is providing instruction in one of the reading strategies or instructional programs that support strategy instruction, the teacher does not demonstrate a thorough understanding of the content. For example, students may not be provided with an in-depth, comprehensive understanding of the strategy and/or program and the teacher, while able to answer basic questions, might not be able to thoroughly respond to more complex questions on the instructional content. As another example, the teacher may be providing comprehensive instruction in the strategy but may not be providing instruction in small steps or segments appropriate for developing student understanding.
3. Instructional content is comprised of instruction in reading strategies (e.g., vocabulary, word-identification, self-questioning, visual imagery, paraphrasing, and inferencing) and other instructional programs that support strategy instruction (ACHIEVE Skills, SCORE Skills, Talking Together, Possible Selves). The teacher demonstrates a strong understanding and knowledge of the content and is able to thoroughly respond to student questions. Further, instruction in the strategy is divided into small steps or segments to facilitate the development of student understanding in this strategy

Fidelity Scale (Core Principle 2.b: Research-based Methodology)

1. The teacher does not use any of the eight instructional stages of the Xtreme Reading Program;* and the learning activities do not appear to be associated with the program's curriculum. Instruction appears unsystematic and unmethodical.
2. The teacher uses one of the eight instructional stages of the Xtreme Reading Program;* however, the teacher does not demonstrate a thorough understanding of the learning activities associated with the specific instructional stage. Although students are involved in learning activities associated with the specific instructional stage, at times, instruction appears unsystematic.
3. The reading strategy of focus is taught using one of the eight stages of the Xtreme Reading instructional methodology. The teacher engages students in learning activities associated with at least one of the eight instructional stages of the Xtreme Reading Program.* The teacher's implementation of the instructional stage reflects best practices, as outlined by the Xtreme Reading instructional methodology, and instruction is delivered in a systematic manner.

* The eight instructional stages are: Describe, Model, Verbal Practice, Guided Practice, Paired Practice, Independent Practice, Differentiated Practice, Generalization

Fidelity Scale (Core Principle 2.c: Connected, Scaffolded, and Informed Instruction)

1. Instruction is neither connected, scaffolded, nor informative. In almost all instances, the teacher does not show students how new information is related to skills, strategies, or content that they have previously learned or that will be learned in the future. Course and Unit Organizers are rarely used for this purpose. There is little evidence of the teacher providing multiple instructional supports (i.e. modeling, prompts, direct explanations, targeted questions, etc.) to facilitate movement from teacher-mediated to student-mediated instruction. The teacher rarely engages students in discussion regarding their own learning process, learning expectations, and why it is important for students to take control of their own learning.
2. Instruction may be connected, scaffolded, or informative, but it does not reflect all three characteristics. In some cases, the teacher provides a brief explanation of how new information is related to skills, strategies, or content that has been previously learned, as well as to those that will be learned in the future. The teacher uses Course and Unit Organizers to introduce new information but does not engage students to ensure their understanding. The teacher provides students with some instructional supports, but not in a systematic manner to promote movement from teacher-mediated to student-mediated instruction. Occasionally, the teacher engages students to ensure they understand how they are progressing, to inform students of how they can control their own learning and why this is important.
3. Instruction is connected, scaffolded, and informative. The teacher purposefully shows students how new information is related to skills, strategies, or content that has been previously learned, as well as to those that will be learned in the future. Course and Unit Organizers are provided to students to introduce main ideas and to demonstrate how critical information and concepts are related. The teacher provides students with multiple instructional supports (i.e. modeling, prompts, direct explanations, targeted questions, etc.) that promote movement from teacher-mediated to student-mediated instruction. The teacher informs students about how the learning process works and what is expected during instruction. The teacher ensures students understand how they are progressing, how they can control their own learning and why this is important.

Core Principle # 3**Classroom Management**

Classroom management and planning techniques maximize the use of instructional time.

- Expectations for all activities and transitions between activities are explained, taught, and reinforced throughout instruction.
- Classroom routines are established early, and students demonstrate familiarity and comfort with these routines.
- Lessons are clearly structured, and all instructional time is used for instruction.
- Interactive learning experiences ensure that students practice, master, integrate, and generalize critical skills.

Fidelity Scale

1. There is little or no evidence of established classroom management techniques. Students do not seem familiar or comfortable with classroom routines. Instructional time is lost due to disorganized transitions between activities and to disciplinary matters. This could take the shape of disorganized, poorly structured instructional activities. As another example, the teacher may not articulate explicit expectations for activities and transitions.
2. Although classroom management techniques appear to be in place, they do not always serve to maximize instruction. At times, students demonstrate a familiarity and comfort with classroom routines. For example, teacher expectations may be articulated for some activities, but are not always reinforced throughout instruction. Some lessons are clearly structured and most instructional time is used for instruction. As another example, interactive learning experiences allow students to practice, master, integrate, and generalize critical skills, but at times students need to be redirected to stay on-task and on-topic.
3. Classroom management techniques maximize the use of instructional time. Students demonstrate a familiarity and comfort with classroom routines and remain focused throughout the instructional period. Instruction fitting this category could take the form of clear and explicit teacher expectations for all activities and transitions between activities that are reinforced throughout the instruction. As another example, lessons are clearly structured and all instructional time is used for instruction. Interactive learning experiences ensure that students practice, master, integrate, and generalize critical skills.

Core Principle # 4**High Student Motivation and Engagement**

Instruction reflects high student motivation and engagement.

- *Student Engagement:* Engagement is maintained in the classroom through activities that enable students to focus attention on critical learning outcomes. Instruction demands a high degree of student attention and response, and expectations are set high for student work. Instruction is interactive and appropriately paced to maintain student attention.
- *Student Motivation:* Motivation is achieved by providing students with a real purpose for improving their literacy skills and by linking learning to their personal goals. In addition, interesting novels are used to motivate students to engage in reading activities.

Fidelity Scale

1. There is little or no evidence of student engagement in classroom activities, and there are few if any opportunities for active learning. For example, the pacing of instruction does not maintain student engagement; students demonstrate boredom and/or frustration regarding the content being taught. As another example, teacher expectations for quality student work and performance appear to be low.

The teacher does not provide students with a real purpose for improving their literacy skills and engaging in the lesson activities. For example, there is little evidence to suggest students are provided with interesting novels to read while engaging in reading activities.

2. During some activities, student engagement is maintained through activities that require a high degree of student attention and response; however, not all students are engaged at all times. For example, the pacing of instruction appears appropriate for some students, but others demonstrate boredom and/or frustration with the content being taught.

At times, the teacher provides students with a purpose for improving their literacy skills, but this purpose is not always clearly relevant, or clearly linked to students' personal goals. It appears that students have access to novels in the classroom, but it is unclear the extent to which these reading materials are used to engage students in reading activities.

3. Student engagement is maintained in the classroom through activities that enable students to focus attention on critical learning outcomes. Instruction demands a high degree of student attention and response, and expectations are set for high-quality student work. Instruction is interactive and appropriately paced to maintain student attention.

The teacher facilitates student motivation by providing students with a real purpose for improving their literacy skills and by linking learning to their personal goals. Additionally, interesting novels are used to motivate students to engage in reading activities.

Appendix E

Technical Notes for Early Impact Findings

This appendix provides two sets of additional technical notes that accompany the impact findings presented in Chapter 5. The first section presents tables that show the effect of covariates on the core impact findings for the full sample of 34 schools and for the groups of schools using each of the two supplemental literacy programs. These tables also present the standard errors (“S.E.” in the tables) and 95 percent confidence intervals for the adjusted and unadjusted impacts. The second section addresses the issues related to multiple hypothesis tests of impacts on multiple reading behavior measures. Specifically, it presents the findings from the qualifying tests that were performed to assess the robustness of the statistical significance of the impacts on the three reading behavior measures examined in Chapter 5.

Adjusted and Unadjusted Impact Estimates

The early impacts presented in Chapter 5 of this report were estimated using regression adjustments for random differences between the ERO and non-ERO groups in their pretest scores and whether a student was overage for the ninth grade. The first two tables in this appendix provide both regression-adjusted and unadjusted impacts. These tables also include other information that may be useful to those who may wish to include these early impacts in meta-analyses. Note that random assignment of students to the ERO and non-ERO groups occurred within each high school (that is, random assignment was “blocked” by school). Because of differences across schools (blocks) in the number of students eligible and appropriate for the ERO programs, the ratio of ERO group members to non-ERO group members in each site varies from 1.22 to 2.0. Thus, all the impact estimates presented in this report include controls for each block to account for random differences between the ERO and non-ERO groups that may be associated with differences in the random assignment ratios. The assessment of sensitivity to other regression adjustments presented in the appendix reflects potential differences in impact estimates that also controls for the blocking of random assignment by school.

Appendix Table E.1 is the counterpart to Tables 5.1 and 5.2 and shows adjusted and unadjusted impacts on reading achievement for all 34 schools in the study and for the groups of schools using each of the two ERO programs. Appendix Table E.2 is the counterpart to Tables 5.3 and 5.4 and shows adjusted and unadjusted impacts on reading behavior measures.¹

¹Results from the regression-adjusted impact analyses are presented in the columns under “Regression-Based Impact Estimates,” and results from the unadjusted impact analyses are presented in the columns under “Mean Differences Adjusted for Blocking Only.”

The Enhanced Reading Opportunities Study
Appendix Table E.1
Impacts on Reading Achievement,
Cohort 1 Follow-Up Respondent Sample

Outcome	Mean Differences Adjusting for Blocking Only				Regression-Based Impact Estimates								
	ERO Group (S.D.)	Non-ERO Group (S.D.)	95% Confidence Interval (S.E.)	P-Value for the Difference	Estimated Impact (S.E.)	95% Confidence Interval (S.E.)	P-Value of Estimated Impact	Impact (S.E.)	95% Confidence Interval				
<u>All schools</u>													
Reading comprehension	90.15	89.53	0.61	-0.20	1.43	0.140	0.90	0.15	1.66	0.019	0.09	0.01	0.16
Average standard score	(10.46)	(10.46)	(0.41)				(0.38)				(0.04)		
Reading vocabulary	93.45	93.43	0.01	-0.80	0.83	0.976	0.28	-0.48	1.04	0.472	0.03	-0.05	0.10
Average standard score	(10.25)	(10.25)	(0.42)				(0.39)				(0.04)		
Sample size	1,408	1,005											
<u>Reading Apprenticeship schools</u>													
Reading comprehension	89.79	88.94	0.85	-0.34	2.03	0.160	0.94	-0.17	2.04	0.097	0.09	-0.02	0.20
Average standard score	(10.35)	(10.35)	(0.60)				(0.56)				(0.05)		
Reading vocabulary	93.25	92.85	0.40	-0.78	1.57	0.508	0.48	-0.62	1.57	0.393	0.05	-0.06	0.15
Average standard score	(10.18)	(10.18)	(0.60)				(0.56)				(0.05)		
Sample size	686	454											
<u>Xtreme Reading schools</u>													
Reading comprehension	90.49	90.08	0.41	-0.71	1.53	0.476	0.89	-0.14	1.93	0.090	0.09	-0.01	0.18
Average standard score	(10.56)	(10.56)	(0.57)				(0.53)				(0.05)		
Reading vocabulary	93.64	93.96	-0.32	-1.45	0.81	0.576	0.11	-0.96	1.17	0.846	0.01	-0.09	0.11
Average standard score	(10.32)	(10.32)	(0.58)				(0.54)				(0.05)		
Sample size	722	551											

(continued)

Appendix Table E.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 2006 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

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 = 10.458; reading vocabulary = 10.505).

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.

The Enhanced Reading Opportunities Study
Appendix Table E.2
Impacts on Reading Behaviors,
Cohort 1 Follow-Up Respondent Sample

Outcome	Mean Differences Adjusting for Blocking Only				Regression-Based Impact Estimates				
	ERO Group (S.D.)	Non-ERO Group (S.D.)	Difference (S.E.)	95% Confidence Interval	P-Value for the Difference	Estimated Impact (S.E.)	95% Confidence Interval	P-Value of Estimated Impact	95% Confidence Interval
All schools									
Amount of school-related reading (prior month occurrences)	44.17 (44.70)	43.45 (44.70)	0.73 (1.81)	-2.83 4.28	0.689	0.78 (1.82)	-2.79 4.34	0.669	0.02 (0.04)
Amount of non-school-related reading (prior month occurrences)	27.26 (30.38)	25.90 (30.38)	1.36 (1.28)	-1.15 3.87	0.289	1.29 (1.28)	-1.23 3.80	0.315	0.04 (0.04)
Use of reflective reading strategies (prior month occurrences)	2.61 (0.64)	2.62 (0.64)	0.00 (0.03)	-0.06 0.05	0.853	-0.01 (0.03)	-0.06 0.05	0.849	-0.01 (0.04)
Sample size	1,410	1,002							
Reading Apprenticeship schools									
Amount of school-related reading (prior month occurrences)	43.85 (43.98)	48.40 (43.98)	-4.55 (2.72)	-9.88 0.77	0.094	-4.48 (2.72)	-9.81 0.86	0.100	-0.10 (0.06)
Amount of non-school-related reading (prior month occurrences)	26.80 (28.64)	27.57 (28.64)	-0.77 (1.87)	-4.44 2.90	0.681	-0.79 (1.87)	-4.47 2.88	0.672	-0.02 (0.06)
Use of reflective reading strategies (prior month occurrences)	2.64 (0.63)	2.66 (0.63)	-0.02 (0.04)	-0.10 0.06	0.607	-0.02 (0.04)	-0.10 0.06	0.600	-0.03 (0.06)
Sample size	689	455							
Xtreme Reading schools									
Amount of school-related reading (prior month occurrences)	44.48 (45.40)	39.15 (45.40)	5.33 (2.43)	0.57 10.09	0.028	5.31 (2.43)	0.54 10.09	0.029	0.12 (0.06)
Amount of non-school-related reading (prior month occurrences)	27.70 (31.96)	24.47 (31.96)	3.22 (1.76)	-0.22 6.67	0.067	3.07 (1.76)	-0.38 6.52	0.081	0.10 (0.06)
Use of reflective reading strategies (prior month occurrences)	2.59 (0.66)	2.58 (0.66)	0.01 (0.04)	-0.06 0.08	0.824	0.01 (0.04)	-0.06 0.08	0.779	0.02 (0.06)
Sample size	721	547							

(continued)

Appendix Table E.2 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 43.867; non-school-related reading standard deviation = 31.834; use of reading strategies standard deviation = 0.670).

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.

For each of the above measures, data are missing for no more than 5 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

Addressing Risks Associates with Multiple Hypothesis Tests

In Chapter 5, statistical significance is indicated in the tables by an asterisk (*) when the p-value of the impact estimate is less than or equal to 0.05 (5 percent). As discussed in Chapter 2, however, when making judgments about statistical significance, it is important to recognize potential problems associated with conducting multiple hypothesis tests. Specifically, it is important to minimize the risk that conclusions from the study could be based on false positive results (also known as Type I errors) while simultaneously limiting the risk that important results may be neglected due to false negative results (also known as Type II errors). In other words, the analysis should avoid concluding that an impact estimate is statistically significant when, in fact, there is no true impact. Likewise the analysis should not be so conservative with respect to producing false positives that it unduly increases the likelihood of missing true impacts when they exist (that is, of producing false negatives).

As the number of hypothesis tests increases, the probability of finding a statistically significant impact estimate when there is no true impact may also increase. One could dramatically reduce this risk by making the standard for statistical significance much more stringent, for example, by setting the p-value to less than or equal to 0.001. Making the standard too stringent, however, will increase the likelihood that one would judge an impact estimate to be not statistically significant when, in fact, it represents a true impact. The approach adopted for this project provides a framework that aspires for an *acceptable balance* between the risks of making Type I and Type II errors.

The impact analysis conducted for this report includes two sets of safeguards aimed at attenuating the risk of drawing inappropriate conclusions about program effectiveness on the basis of multiple hypothesis tests. The first safeguard is to identify a *parsimonious* list of outcome measures and subgroups and then to prioritize among these to specify the primary and secondary hypothesis tests that would be used to make judgments about the overall effectiveness of the ERO programs. The shorter this list, the fewer the number of hypothesis tests and, thus, the less exposed the analysis will be to “spurious statistical significance” as a result of having tested multiple hypotheses.

The second safeguard uses composite statistical tests to “qualify” or call into question multiple hypothesis tests that are statistically significant individually but that may be due to chance. These composite tests are referred to as “qualifying tests.”

Specifying Primary and Secondary Hypothesis Tests

The primary evidence of overall ERO program effectiveness for this report will be reflected by estimates of program impacts on reading comprehension test scores (expressed in

standard score values) for the full study sample and for each of the two ERO programs being evaluated. Anchoring the study's early conclusions in a limited set of outcomes minimizes the risk of relying on a large number of impact estimates, some of which may be statistically significant only by chance. As noted above, student reading comprehension skills constitute the primary target of the ERO interventions and the primary outcome of interest for the first year of the study. Also, the study was designed to provide minimum detectable effect sizes for each ERO subgroup that may be considered policy relevant. Thus, the primary confirmatory hypotheses for the report focus on the overall and program-specific impacts on reading comprehension test scores.

Vocabulary knowledge and student reading behaviors, while targets of the interventions and important to students' literacy development, are considered secondary indicators of program effectiveness. Similarly, subgroups of students (for example, those with higher or lower baseline test scores) and subgroups of schools (for example, those that were able to operate for longer or shorter periods of time during the first year) provide useful information about the relative impact of supplemental literacy programs, but they too are considered secondary indicators of effectiveness in this report.

Composite Qualifying Statistical Tests

A second set of safeguards against risks associated with multiple hypothesis tests involves the use of composite qualifying statistical tests that provide further context for interpreting the robustness of individual impact estimates and their statistical significance.² These statistical tests are applied in cases where impacts are estimated for more than one outcome in a given measurement domain (for example, the three survey measures that attempt to capture students' reading behaviors) or for subgroups of the full study sample. In general, these qualifying statistical tests estimate impacts on composite indices that encompass all the measures in a given domain or estimate the overall variation in impacts across subgroups. If the results of these tests are not statistically significant, this indicates that the statistical significance of the associated individual impact estimates *may have* occurred by chance. In these cases, the discussion of the impacts should include cautions or qualifiers about the robustness of the individual findings.³

²Measurement of overall effects has its roots in the literature on meta-analysis (see O'Brien, 1984; Logan and Tamhane, 2003; and Hedges and Olkin, 1985). For a discussion of qualifying statistical tests to account for the risk of Type I error, see Duflo, Glennerster, and Kremer (2007). Other applications of these approaches are discussed in Kling and Liebman (2004) and Kling, Liebman, and Katz (2007).

³Alternative strategies that involve (1) adjusting significance levels (through Bonferroni methods) or (2) adjusting significance thresholds (through Benjamini and Hochberg methods) are overly conservative with respect to making Type I errors and can thereby greatly increase the likelihood of making Type II errors. There are two reasons for this. First, these methods treat all hypotheses as though they were independent of each other. Hence, each hypothesis is treated as representing an independent opportunity to make a Type I error. How-

(continued)

To test the robustness of the statistical significance of impact estimates for multiple outcomes within a measurement domain (in this case, the three reading behavior measures), the study uses a single composite index consisting of the average of the standardized values for each outcome.⁴ Then the estimated impact on this composite measure is calculated for the full study sample. If this qualifying test shows that the composite impact estimate is not statistically significant (its p-value is greater than 0.05), then one concludes that statistically significant impacts for the component outcomes could be due to chance and should be interpreted cautiously.

Specifically, the analysis took the following steps in creating a composite index and assessing impacts on reading behaviors.⁵ First, z-scores were created for each reading behavior outcome by subtracting the non-ERO group mean and dividing by the non-ERO group standard deviation. Thus, each component of the index has a mean of zero and a standard deviation of one for the non-ERO group. The z-scores from each component were averaged to obtain the index which was then included in the standard impact estimation model. If the estimated impact for the composite index is not statistically significant, then the statistical significance of impact estimates for the component measures may have occurred by chance and the finding should be interpreted cautiously. In other words, the report qualifies or calls into question a statistically significant individual impact estimate by suggesting that it may have occurred by chance.

To test the robustness of the statistical significance of impact estimates for subgroups of students or schools, a composite F-test is used to assess whether the variation in impacts across all student or school subgroups is statistically significant. For example, the analysis examines impacts for three sets of student subgroups: those defined by baseline reading test scores (comprising three subgroups); those defined by whether a student was overage for the start of ninth grade (comprising two subgroups); and those defined by whether a student's family spoke a language other than English at home (comprising two subgroups). The composite qualifying test for these analyses assesses whether variation in estimated impacts across these seven sub-

ever, many impact estimates in an evaluation study are correlated with each other and thus do not represent independent opportunities to make Type I errors. In the extreme, for example, if all measures were perfectly correlated, there is only one opportunity to make a Type I error even though there are many outcome measures and, thus, many statistical hypothesis tests. The above methods assume, however, that the number of opportunities to make a Type I error equals the number of hypothesis tests conducted. To the degree that hypothesis tests are correlated with each other, these methods overcompensate (often by a lot) for the risks of Type I error in multiple hypothesis tests. A second source of conservatism with respect to Type I error is the fact that the above methods assume that all null hypotheses may be true. As a result, they consider the potential number of false positives to equal the total number of hypothesis tests conducted. However, the actual number of potential false positives equals the total number of true null hypotheses, not the total number of hypotheses tested. This is because only true null hypotheses can produce false positives. Hence, the methods overcompensate for the number of hypotheses tested.

⁴See Duflo, Glennerster, and Kremer (2007).

⁵The discussion and method presented here draw from Kling, Liebman, and Katz (2007).

groups accounts for a statistically significant level of unexplained variance in the test score or other outcome being examined. In other words, the test assesses whether the change in the F-statistic from the core impact regression to the impact regression with the subgroup interaction terms is statistically significant (its p-value is less than or equal to 0.05). If the change in unexplained variance due to the subgroup impact interactions is not statistically significant, then the statistical significance of impact estimates for the component subgroups may have occurred by chance and the findings should be interpreted cautiously.

Finally, the analysis includes qualifying statistical tests to assess the statistical significance of the difference in impacts between the subgroups of students or schools. If these qualifying tests show that the difference in impacts across subgroups is not statistically significant (p-value is greater than 0.05), then one concludes that statistically significant impacts for individual subgroups could be due to chance and should be interpreted cautiously.⁶ For example, suppose the findings indicate that impacts on reading comprehension for one group of participating high schools are positive and statistically significant while the result for a second group of schools is also positive but is not statistically significant. If the difference in impacts between the two groups of schools is not statistically significant, one should be especially cautious about concluding that the ERO programs were more effective for some schools than for others.

Appendix Table E.3 displays the results of the composite qualifying statistical tests for the three reading behavior measures discussed in Chapter 5. As discussed above, the composite index was created by averaging the standardized values of the three reading behaviors outcomes: amount of school-related reading, amount of non-school-related reading, and use of reflective reading strategies. Appendix Table E.3 shows results for the full sample of all schools, for each of the two ERO programs separately, and for the various subgroups that are discussed in Chapter 5. None of the estimated impacts on the composite index is statistically significant. Thus, readers should exercise caution in interpreting statistically significant impacts for the individual components of the composite index, since these may be due to chance.

Appendix Table E.3 also includes the results of the composite qualifying statistical test of the robustness of statistical significance of the *difference in impacts* across subgroups of students or schools. It shows that even though none of the impact estimates themselves is statistically significant, the difference in impacts is statistically significant for three sets of subgroups: those for each of the two ERO programs, those defined by language spoken at home, and those defined by first-year implementation issues. Thus, the difference in impacts should be interpreted cautiously, given that the ERO programs did not produce statistically significant impacts on the composite index for the full sample or for any of the subgroups.

⁶Note that one conducts qualifying statistical tests using the composite index when assessing the robustness of impacts for multiple measures across multiple subgroups of the study sample.

The Enhanced Reading Opportunities Study
Appendix Table E.3
Impacts on Reading Behaviors Composite Index,
for the Full Study Sample and Subgroups

Subgroup	Estimated Impact	P-Value for Estimated Impact
<u>All schools</u>	0.02	0.529
<u>Programs</u>		
Reading Apprenticeship schools	-0.05	0.311
Xtreme Reading schools	0.08	0.065
<i>Difference in impacts</i>	-0.12 *	0.046
<u>Baseline comprehension performance</u>		
6.0-7.0 grade equivalent	0.02	0.668
5.0-5.9 grade equivalent	0.05	0.365
4.0-4.9 grade equivalent	0.00	0.958
<i>Difference in impacts, 6.0-7.0 minus 5.0-5.9</i>	-0.03	0.712
<i>Difference in impacts, 6.0-7.0 minus 4.0-4.9</i>	0.03	0.727
<u>Overage for grade^a</u>		
Student is overage for grade	0.05	0.481
Student is not overage for grade	0.01	0.786
<i>Difference in impacts</i>	0.04	0.626
<u>Language spoken at home</u>		
Students from multilingual families	0.08	0.086
Students from English-only families	-0.05	0.268
<i>Difference in impacts</i>	0.13 *	0.045
<u>First-year Implementation issues</u>		
Fewer implementation issues	0.08	0.058
More implementation issues	-0.04	0.360
<i>Difference in impacts</i>	0.12 *	0.046

(continued)

Appendix Table E.3 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The reading behaviors composite index is the average of the standardized values of the three reading behavior measures: amount of school-related reading, amount of non-school-related reading, and use of reflective reading strategies. The values were standardized using the non-ERO group mean and standard deviation.

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.

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 differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

Appendix F

Early Impact Estimates Weighted for Nonresponse

As discussed in Appendix B, the response analysis revealed several differences between students who completed the follow-up test and those who did not. Most notably, there were differences in response rates between the ERO group and the non-ERO group and there was variation across the participating high schools. In addition, nonrespondents were more likely to be overage for the ninth grade and to have lower pretest scores. As a result, students with these characteristics are underrepresented in the sample used to estimate impacts. The over- or underrepresentation of students with certain characteristics in the impact analysis sample may lead to findings that cannot be generalized to the original sample.

This appendix assesses the sensitivity of the impact estimates to the over- or underrepresentation of key baseline characteristics in the impact analysis sample. Specifically, it examines impact estimates that are weighted to account for differential response rates between the ERO and non-ERO groups and across high schools and that impact estimates are associated with being overage for grade and with differences in baseline test scores. Sampling weights were constructed using multiple regressions in which response rates were predicted based on a student's baseline test score and an indicator of whether the student was overage for the ninth grade. Separate regressions were estimated for each high school and for the ERO students and non-ERO students within each school. The sampling weights were constructed as the inverse of the predicted response rate for each student in the full study sample.

These sampling weights ensure that each high school and the ERO and non-ERO groups within each high school can be represented in the impact analysis in the same proportion as they are in the full study sample. They also ensure that the distribution of overage-for-grade baseline tests scores in the impact sample is equivalent to their representation in the full sample.

Appendix Table F.1 displays the weighted impact estimates for reading achievement for all 34 high schools and for the schools using each of the two supplemental reading programs. It shows that, together, the ERO programs produced a statistically significant weighted impact on reading comprehension of 1.0 standard score (a 0.09 effect size). This is slightly larger than the estimated impact for the respondent sample (0.9 standard score point). As with the results for the respondent sample, neither program alone produced a statistically significant weighted impact on reading comprehension test scores, although the magnitude of the weighted impact estimates are the same as the impact for the full sample. Appendix Table F.1 also shows that the ERO programs did not have a statistically significant weighted impact on vocabulary test scores.

Appendix Table F.2 displays the weighted impacts on the reading behavior measures. These results are nearly the same as those estimated with the respondent sample and displayed in Tables 5.3 and 5.4.

In summary, differences between students who completed the follow-up test and survey and those who did not do not appear to change the underlying pattern of impacts on test scores or reading behaviors.

The Enhanced Reading Opportunities Study

Appendix Table F.1

Impacts on Reading Achievement Weighted by School Response Rate,
Cohort 1 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	90.0	89.0	1.0 *	0.09 *	0.008
<i>Corresponding grade equivalent</i>	6.1	5.9			
<i>Corresponding percentile</i>	25	23			
Reading vocabulary					
Average standard score	93.3	93.0	0.3	0.03	0.396
<i>Corresponding grade equivalent</i>	7.7	7.7			
<i>Corresponding percentile</i>	32	31			
Sample size	1,408	1,005			
<u>Reading Apprenticeship schools</u>					
Reading comprehension					
Average standard score	89.6	88.5	1.1	0.09	0.055
<i>Corresponding grade equivalent</i>	6.0	5.8			
<i>Corresponding percentile</i>	24	22			
Reading vocabulary					
Average standard score	93.0	92.5	0.5	0.04	0.381
<i>Corresponding grade equivalent</i>	7.7	7.7			
<i>Corresponding percentile</i>	31	30			
Sample size	686	454			
<u>Xtreme Reading schools</u>					
Reading comprehension					
Average standard score	90.4	89.4	1.0	0.08	0.062
<i>Corresponding grade equivalent</i>	6.2	6.0			
<i>Corresponding percentile</i>	25	24			
Reading vocabulary					
Average standard score	93.6	93.4	0.2	0.02	0.740
<i>Corresponding grade equivalent</i>	7.8	7.7			
<i>Corresponding percentile</i>	32	32			
Sample size	722	551			

(continued)

Appendix Table F.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 2006 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

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 = 11.599; reading vocabulary = 11.654).

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.

The Enhanced Reading Opportunities Study

Appendix Table F.2

**Impacts on Reading Behaviors Weighted by School Response Rate,
Cohort 1 Follow-Up Respondent Sample**

Outcome	ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>All schools</u>					
Amount of school-related reading (prior month occurrences)	44.53	43.25	1.28	0.03	0.485
Amount of non-school-related reading (prior month occurrences)	27.63	26.11	1.52	0.04	0.242
Use of reflective reading strategies in class (4-point scale)	2.62	2.62	0.00	0.00	0.911
Sample size	1,410	1,002			
<u>Reading Apprenticeship schools</u>					
Amount of school-related reading (prior month occurrences)	44.11	47.82	-3.71	-0.08	0.176
Amount of non-school-related reading (prior month occurrences)	27.02	27.69	-0.66	-0.02	0.726
Use of reflective reading strategies in class (4-point scale)	2.65	2.65	-0.01	-0.01	0.857
Sample size	689	455			
<u>Xtreme Reading schools</u>					
Amount of school-related reading (prior month occurrences)	44.92	39.33	5.59 *	0.11 *	0.023
Amount of non-school-related reading (prior month occurrences)	28.20	24.79	3.41	0.10	0.057
Use of reflective reading strategies in class (4-point scale)	2.59	2.59	0.00	0.00	0.923
Sample size	721	547			

(continued)

Appendix Table F.2 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 48.992; non-school-related reading standard deviation = 35.864; use of reading strategies standard deviation = 0.749).

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.

For each of the above measures, data are missing for no more than 5 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

Appendix G

**Early Impacts on Supplementary Measures of
Reading Achievement and Behaviors**

In an effort to understand more about the extent and nature of ERO program impacts on student outcomes, the ERO study team performed secondary impact analyses. These analyses fall into two categories. First, the supplemental analyses explore additional measures from the ERO follow-up student survey. These measures were created to complement the reading behaviors measures discussed in the report. They contribute to a more detailed picture of how the program changed or did not change students' attitudes toward reading and their behavior in school. Second, the study team analyzed the impact of the ERO program on the percentage of students who were less than two years behind grade level in reading by the end of the school year. Given that students needed to be at least two years below grade level in reading to be eligible for the program, those students who have attained reading levels above this cutoff have succeeded in moving beyond the scope of the program during the school year.

Impacts on Students' Attitudes and Perceptions of Reading and School

As discussed in Appendix A, the ERO follow-up student survey included a variety of questions related to students' attitudes and perceptions of reading and school. Beyond the three reading behaviors measures discussed in the report, several secondary measures were explored, including students' attitudes toward literacy, whether or not they believe that reading is connected to learning, how easy they feel it is to read different types of texts for school, their persistence in successfully completing schoolwork, whether or not they display negative school behaviors such as cutting class or disobeying school rules, and what their educational aspirations are.

These measures are not included in the report because they were less directly related to ERO program goals or less likely to display short-term impacts. Appendix Table G.1 shows the impact findings for each of these six measures. The only construct showing statistically significant positive impacts is the measure of positive literacy attitudes. It quantifies whether students enjoy reading and writing and consider them useful activities for learning new ideas and expressing themselves. There are also statistically significant impacts on this measure for students in the Xtreme Reading schools, suggesting that this specific program had a small, positive effect on students' attitudes toward reading and writing.

The Impacts on the Percentage of Students No Longer Eligible for the ERO Programs

Both Reading Apprenticeship Academic Literacy and Xtreme Reading attempt to accelerate literacy learning through their instructional programs to help struggling students attain the reading skill levels needed to succeed in high school classes. One way of measuring the impact of the ERO program is to look at whether more ERO students are bridging this gap in skills

during their first year of high school students who did not participate in ERO. To answer this question, the study team analyzed the program impact on the percentage of students who were less than two years behind grade level in reading comprehension by the end of the school year, and, therefore, were no longer eligible for the program. The percentage of ERO program students whose follow-up GRADE standard score for reading comprehension was a 98 or above and whose corresponding grade equivalent was at least 8.2 were compared with the percentage of non-ERO students who scored at or above this level on the GRADE follow-up test. As shown in Appendix Table G.2, the ERO program impacts for the entire sample and for each of the programs are small and are not statistically significant at the 5 percent level.

The Enhanced Reading Opportunities Study
Appendix Table G.1
Impacts on Attitudes and Perceptions of Reading and School,
Cohort 1 Follow-Up Respondent Sample

Outcome	ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>All schools</u>					
Positive Literacy Attitudes (4-point scale)	2.47	2.42	0.05 *	0.08 *	0.042
Reading to Learn (4-point scale)	2.63	2.61	0.02	0.04	0.370
Ease of Reading (4-point scale)	2.88	2.91	-0.03	-0.05	0.242
Persistence on School Work (4-point scale)	2.76	2.78	-0.03	-0.04	0.305
Negative School Behavior (4-point scale)	1.09	1.12	-0.03	-0.02	0.570
Educational Aspiration (binary)	0.64	0.64	-0.01	-0.01	0.752
Sample size	1,410	1,002			
<u>Reading Apprenticeship schools</u>					
Positive Literacy Attitudes (4-point scale)	2.50	2.47	0.03	0.05	0.432
Reading to Learn (4-point scale)	2.67	2.63	0.04	0.06	0.259
Ease of Reading (4-point scale)	2.86	2.90	-0.04	-0.07	0.232
Persistence on School Work (4-point scale)	2.75	2.82	-0.06	-0.10	0.101
Negative School Behavior (4-point scale)	1.11	1.08	0.02	0.02	0.773
Educational Aspiration (binary)	0.62	0.66	-0.03	-0.06	0.300
Sample size	689	455			
<u>Xtreme Reading schools</u>					
Positive Literacy Attitudes (4-point scale)	2.45	2.37	0.07 *	0.11 *	0.037
Reading to Learn (4-point scale)	2.60	2.59	0.01	0.01	0.869
Ease of Reading (4-point scale)	2.91	2.93	-0.02	-0.03	0.600
Persistence on School Work (4-point scale)	2.76	2.75	0.01	0.01	0.876
Negative School Behavior (4-point scale)	1.08	1.16	-0.08	-0.06	0.257
Educational Aspiration (binary)	0.65	0.63	0.02	0.03	0.557
Sample size	721	547			

(continued)

Appendix Table G.1 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (Positive Literacy Attitudes standard deviation = 0.650; Reading to Learn standard deviation = 0.668; Ease of Reading standard deviation = 0.510; Persistence on School Work standard deviation = 0.636; Negative School Behavior standard deviation = 1.205; Educational Aspiration standard deviation = 0.480).

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.

For each of the above measures, data are missing for no more than 14 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

The Enhanced Reading Opportunities Study

Appendix Table G.2

**Impacts on Percentage of Students No Longer Eligible for Program,
Cohort 1 Follow-Up Respondent Sample**

Outcome	ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>All schools</u>					
No longer eligible for program ^a (%)	23.93	21.42	2.52	0.06	0.125
Sample size	1,408	1,005			
<u>Reading Apprenticeship schools</u>					
No longer eligible for program (%)	22.74	20.65	2.09	0.05	0.374
Sample size	686	454			
<u>Xtreme Reading schools</u>					
No longer eligible for program (%)	25.07	22.11	2.96	0.07	0.197
Sample size	722	551			

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 2006 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (standard deviation = 41.705).

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.

^aStudents with scores on the GRADE pretest between two and five years below grade level were eligible for the program. Students are considered no longer eligible for the program if their score on the follow-up GRADE assessment is equal to or higher than a standard score of 98 (corresponding grade equivalent of 8.2), suggesting that the student is now less than two years behind grade level.

Appendix H

Early Impacts for Student Subgroups

While all students in the study sample had baseline reading comprehension skills from the fourth- through seventh-grade level at the start of ninth grade, the ERO study sample includes a diverse population of students. With this diversity in mind, the ERO evaluation was designed to allow for the estimation of impacts for key subgroups of students who face especially challenging barriers to literacy development and overall school performance in high school. For example, prior research has shown that especially low literacy levels, evidence of failure in prior grades, and having English as a second language are powerful predictors of school success.¹

This appendix examines variation in ERO program impacts for subgroups of students defined by their baseline reading comprehension test scores, whether they were overage for the ninth grade, and whether a language other than English was spoken in their homes. As reported in Chapter 2 (see Table 2.4), 36 percent of the study sample had baseline test scores that indicate reading levels that were four to five years below grade level at the start of ninth grade, and another 28 percent were reading from three to four years below grade level. Also, over a quarter of the students in the study sample were overage for the ninth grade, which is used to indicate that a student was retained in a prior grade.² Over 45 percent of the students in the sample lived in households where a language other than English was spoken.

- **Differences in impacts across subgroups of students with different baseline reading comprehension test scores are not statistically significant.**

Appendix Tables H.1 and H.2 correspond to the top panel of Table 5.5 and present impact findings for the subgroups of students defined by their baseline reading comprehension test scores. Appendix Table H.1 indicates that the ERO program produced positive and statistically significant impacts on vocabulary test scores for students whose scores were from two to three years below grade level. Although the impact on vocabulary test scores for this group is statistically significant, the difference between this impact and the impacts for each of the other two subgroups is not statistically significant. Appendix Table H.2 shows that the ERO programs did not produce statistically significant impacts on any of the three measures of reading behaviors for any of the three subgroups defined by baseline test scores.

- **Differences in impacts across subgroups of students who were overage for the ninth grade or not overage for the ninth grade are not statistically significant.**

Appendix Tables H.3 and H.4 correspond to the middle panel of Table 5.5 and present impact findings for the subgroups of students defined by whether they were overage for the

¹Roderick (1993); Fine (1988).

²National Center for Education Statistics (1990).

ninth grade and likely to have been retained in a prior grade. Appendix Table H.3 indicates that the ERO program produced positive and statistically significant impacts on reading comprehension test scores for these students who were overage for grade. Although the impact on reading comprehension test scores for this group is statistically significant, the difference between this impact and the impact for students who were not overage for grade is not statistically significant. Appendix Table H.4 shows that the ERO programs did not produce statistically significant impacts on any of the three measures of reading behaviors for either of the subgroups defined by whether they were overage for grade.

- **Differences in impacts across subgroups of students from multilingual families and those from English-only families are not statistically significant.**

Appendix Tables H.5 and H.6 correspond to the bottom panel of Table 5.5 and present impact findings for the subgroups of students defined by whether a language other than English was spoken in their homes. Appendix Table H.5 indicates that the ERO program produced positive and statistically significant impacts on reading comprehension test scores for students from multilingual families. Although the impact on reading comprehension test scores for this group is statistically significant, the difference between this impact and the impacts for students from English-only families is not statistically significant.

Although Appendix Table H.6 shows that the ERO programs produced a positive and statistically significant impact on the amount of non-school-related reading that students reported, this result should be interpreted cautiously. The qualifying tests conducted for this subgroup of students (see Appendix E) indicate that the ERO programs did not produce a statistically significant impact on the composite index that was created to capture the three reading behavior measures.

To further test any impacts on reading comprehension across all three subgroups, a composite qualifying statistical test for the multiple hypothesis tests was conducted. This test indicates that the overall variation in impacts across all these subgroups is not statistically significant (F -statistic = 0.865; p -value = 0.534), further suggesting that any statistical significance found on reading comprehension impacts for specific subgroups should be interpreted cautiously.

The Enhanced Reading Opportunities Study

Appendix Table H.1

Impacts on Reading Achievement,
Cohort 1 Follow-Up Respondent Sample,
by Baseline Reading Comprehension Performance

Outcome	ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>6.0-7.0 grade equivalent (%)</u>					
Reading comprehension					
Average standard score	94.2	93.1	1.0	0.10	0.106
<i>Corresponding grade equivalent</i>	7.2	6.9			
<i>Corresponding percentile</i>	34	32			
Reading vocabulary					
Average standard score	97.8	96.6	1.3 *	0.12 *	0.040
<i>Corresponding grade equivalent</i>	8.6	8.2			
<i>Corresponding percentile</i>	43	39			
Sample size	485	370			
<u>5.0-5.9 grade equivalent (%)</u>					
Reading comprehension					
Average standard score	90.4	89.6	0.8	0.08	0.274
<i>Corresponding grade equivalent</i>	6.2	6.0			
<i>Corresponding percentile</i>	25	24			
Reading vocabulary					
Average standard score	93.3	94.0	-0.6	-0.06	0.401
<i>Corresponding grade equivalent</i>	7.7	7.8			
<i>Corresponding percentile</i>	32	33			
Sample size	413	267			
<u>4.0-4.9 grade equivalent (%)</u>					
Reading comprehension					
Average standard score	86.1	85.3	0.8	0.08	0.233
<i>Corresponding grade equivalent</i>	5.1	5.0			
<i>Corresponding percentile</i>	17	15			
Reading vocabulary					
Average standard score	89.4	89.6	-0.2	-0.02	0.729
<i>Corresponding grade equivalent</i>	7.1	7.1			
<i>Corresponding percentile</i>	23	23			
Sample size	510	368			

(continued)

Appendix Table H.1 (continued)

Difference in Impacts Between Subgroups	Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>6.0-7.0 minus 5.0-5.9</u>			
Reading comprehension standard score	0.2	0.02	0.821
Reading vocabulary standard score	1.9	0.18	0.051
<u>6.0-7.0 minus 4.0-4.9</u>			
Reading comprehension standard score	0.2	0.02	0.810
Reading vocabulary standard score	1.5	0.14	0.101

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 2006 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

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 = 10.458; reading vocabulary = 10.505).

A two-tailed t-test was applied to the impact estimate and to the difference in impacts. 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.

The Enhanced Reading Opportunities Study
Appendix Table H.2
Impacts on Reading Behaviors,
Cohort 1 Follow-Up Respondent Sample,
by Baseline Reading Comprehension Performance

Outcome	ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>6.0-7.0 grade equivalent (%)</u>					
Amount of school-related reading (prior month occurrences)	43.2	42.3	0.9	0.02	0.760
Amount of non-school-related reading (prior month occurrences)	27.6	24.2	3.4	0.11	0.126
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	-0.06	0.376
Sample size	483	367			
<u>5.0-5.9 grade equivalent (%)</u>					
Amount of school-related reading (prior month occurrences)	45.3	42.6	2.7	0.06	0.430
Amount of non-school-related reading (prior month occurrences)	27.6	26.0	1.6	0.05	0.526
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	0.05	0.471
Sample size	418	267			
<u>4.0-4.9 grade equivalent (%)</u>					
Amount of school-related reading (prior month occurrences)	44.1	44.1	0.0	0.00	0.998
Amount of non-school-related reading (prior month occurrences)	26.7	27.5	-0.8	-0.03	0.691
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	0.00	0.956
Sample size	509	368			

(continued)

Appendix Table H.2 (continued)

Difference in Impacts Between Subgroups	Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>6.0-7.0 minus 5.0-5.9</u>			
Amount of school-related reading	-1.8	-0.04	0.696
Amount of non-school-related reading	1.8	0.06	0.595
Use of reflective reading strategies	-0.1	-0.12	0.257
<u>6.0-7.0 minus 4.0-4.9</u>			
Amount of school-related reading	0.9	0.02	0.832
Amount of non-school-related reading	4.2	0.13	0.164
Use of reflective reading strategies	0.0	-0.06	0.537

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 43.867; non-school-related reading standard deviation = 31.834; use of reading strategies standard deviation = 0.670).

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

For each of the above measures, data are missing for no more than 5 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

The Enhanced Reading Opportunities Study
Appendix Table H.3
Impacts on Reading Achievement,
Cohort 1 Follow-Up Respondent Sample,
by Whether Students Were Overage for Grade

Outcome	ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Overage for grade^a</u>					
Reading comprehension					
Average standard score	88.8	86.8	2.0 *	0.19 *	0.007
<i>Corresponding grade equivalent</i>	5.8	5.3			
<i>Corresponding percentile</i>	22	18			
Reading vocabulary					
Average standard score	91.5	90.6	0.9	0.09	0.221
<i>Corresponding grade equivalent</i>	7.5	7.3			
<i>Corresponding percentile</i>	28	25			
Sample size	395	249			
<u>Not overage for grade</u>					
Reading comprehension					
Average standard score	90.7	90.2	0.5	0.05	0.267
<i>Corresponding grade equivalent</i>	6.2	6.1			
<i>Corresponding percentile</i>	26	25			
Reading vocabulary					
Average standard score	94.2	94.2	0.0	0.00	0.992
<i>Corresponding grade equivalent</i>	7.8	7.8			
<i>Corresponding percentile</i>	33	33			
Sample size	1,013	756			
Difference in Impacts Between Subgroups			Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>Overage minus not overage</u>					
Reading comprehension standard score			1.5	0.14	0.084
Reading vocabulary standard score			1.0	0.09	0.288

(continued)

Appendix Table H.3 (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 2006 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

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 = 10.458; reading vocabulary = 10.505).

A two-tailed t-test was applied to the impact estimate and to the difference in impacts. 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.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

The Enhanced Reading Opportunities Study
Appendix Table H.4
Impacts on Reading Behaviors,
Cohort 1 Follow-Up Respondent Sample,
by Whether Students Were Overage for Grade

Outcome	ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Overage for grade^a</u>					
Amount of school-related reading (prior month occurrences)	43.9	42.2	1.7	0.04	0.667
Amount of non-school-related reading (prior month occurrences)	29.6	26.5	3.2	0.10	0.253
Use of reflective reading strategies (4-point scale)	2.6	2.7	0.0	-0.03	0.676
Sample size	401	250			
<u>Not overage for grade</u>					
Amount of school-related reading (prior month occurrences)	44.3	43.6	0.7	0.02	0.718
Amount of non-school-related reading (prior month occurrences)	26.3	25.7	0.7	0.02	0.647
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	-0.01	0.876
Sample size	1,009	752			
Difference in Impacts Between Subgroups			Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>Overage minus not overage</u>					
Amount of school-related reading			0.9	0.02	0.833
Amount of non-school-related reading			2.5	0.08	0.423
Use of reflective reading strategies			0.0	-0.03	0.777

(continued)

Appendix Table H.4 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 43.867; non-school-related reading standard deviation = 31.834; use of reading strategies standard deviation = 0.670).

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

For each of the above measures, data are missing for no more than 5.5 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

The Enhanced Reading Opportunities Study

Appendix Table H.5

Impacts on Reading Achievement,
Cohort 1 Follow-Up Respondent Sample,
by Language Spoken at Home

Outcome	ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Students from multilingual families</u>					
Reading comprehension					
Average standard score	90.0	88.8	1.2 *	0.12 *	0.027
<i>Corresponding grade equivalent</i>	6.1	5.8			
<i>Corresponding percentile</i>	25	22			
Reading vocabulary					
Average standard score	92.6	91.6	1.0	0.10	0.072
<i>Corresponding grade equivalent</i>	7.7	7.5			
<i>Corresponding percentile</i>	30	28			
Sample size	663	470			
<u>Students from English-only families</u>					
Reading comprehension					
Average standard score	90.3	89.6	0.7	0.07	0.181
<i>Corresponding grade equivalent</i>	6.2	6.0			
<i>Corresponding percentile</i>	25	24			
Reading vocabulary					
Average standard score	94.2	94.6	-0.4	-0.03	0.512
<i>Corresponding grade equivalent</i>	7.8	7.9			
<i>Corresponding percentile</i>	33	34			
Sample size	745	535			
<u>Difference in Impacts Between Subgroups</u>					
			Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>Multilingual minus English-only</u>					
Reading comprehension standard score			0.5	0.05	0.491
Reading vocabulary standard score			1.4	0.13	0.078

(continued)

Appendix Table H.5 (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 2006 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

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 = 10.458; reading vocabulary = 10.505).

A two-tailed t-test was applied to the impact estimate and to the difference in impacts. 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.

The Enhanced Reading Opportunities Study
Appendix Table H.6
Impacts on Reading Behaviors,
Cohort 1 Follow-Up Respondent Sample,
by Language Spoken at Home

Outcome	Non-ERO ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Students from multilingual families</u>					
Amount of school-related reading (prior month occurrences)	45.4	40.4	5.0	0.12	0.052
Amount of non-school-related reading (prior month occurrences)	28.0	24.1	3.9 *	0.12 *	0.031
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	-0.03	0.664
Sample size	660	470			
<u>Students from English-only families</u>					
Amount of school-related reading (prior month occurrences)	43.1	46.8	-3.8	-0.09	0.140
Amount of non-school-related reading (prior month occurrences)	26.6	28.2	-1.6	-0.05	0.387
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	-0.01	0.908
Sample size	750	532			
Difference in Impacts Between Subgroups			Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>Multilingual minus English-only</u>					
Amount of school-related reading			8.8 *	0.20 *	0.015
Amount of non-school-related reading			5.5 *	0.17 *	0.032
Use of reflective reading strategies			0.0	-0.02	0.814

(continued)

Appendix Table H.6 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 43.867; non-school-related reading standard deviation = 31.834; use of reading strategies standard deviation = 0.670).

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

For each of the above measures, data are missing for no more than 4.8 percent of the respondents.

Appendix I

**The Relationship Between Early Impacts and First-Year
Implementation Issues**

This appendix further discusses the impacts for subgroups of the participating high schools that were defined by whether they were able to achieve two implementation milestones during the first year of the study: (1) whether implementation was well aligned or moderately aligned to the respective program models (as defined in Chapter 3) and (2) whether the schools were able to operate for more than seven and a half months (the average for the sample). As discussed in Chapter 5, the 15 schools that were able to reach both these thresholds were deemed to have had a first-year start-up experience that was more in line with the original intent of the program developers than those that did not.

It is important to note that the analyses presented in this appendix are exploratory and are not able to establish causal links between these early implementation challenges and variation in program impacts across the sites.

Appendix Table I.1 is the counterpart to Figure 5.2. It lists the reading comprehension impact estimates of each of the 34 participating high schools in ascending order. It also includes the standard error and 95 percent confidence intervals for these impacts. Four of the 34 schools have statistically significant positive impacts. A composite F-test was used to assess whether the school-level impacts on reading comprehension test scores are statistically equivalent. The F-value is 1.63, and the p-value is 0.013, indicating that the school-to-school variation in impacts is unlikely to have occurred by chance.

Appendix Tables I.2 and I.3 correspond with the top panel of Table 5.6. They display the impacts on reading test scores and reading behaviors, consecutively, for the three groups of schools defined by the fidelity of ERO program implementation during the first year of the study and include the outcome levels for the ERO and non-ERO groups, the impact estimates, p-values, and differences in impacts between the fidelity levels. A statistically significant impact was found for the group of schools whose ERO program implementation was deemed moderately aligned to the program model but was not considered well aligned. The difference in impacts on reading comprehension test scores between the schools deemed moderately aligned and those deemed poorly aligned is statistically significant. Appendix Table I.3 shows that although they are not statistically significant, estimated impacts on the amount of reading students reported are positive for schools with implementation that was either well aligned or moderately aligned and negative for schools with implementation that was poorly aligned.

Appendix Tables I.4 and I.5 correspond with the middle panel of Table 5.6. These tables display the impacts on reading test scores and reading behaviors, consecutively, for the three groups of schools defined by the length of program duration. Appendix Table I.4 shows a statistically significant impact on the reading comprehension estimate for the longest duration schools. The differences in impacts across the three subgroups of sites, however, are not statistically significant. Appendix Table I.5 shows that impacts on the amount of school-related and

non-school-related reading for programs that were able to operate for more than eight months are not statistically significant.

To further test the impacts on reading comprehension for both implementation fidelity and duration, a composite qualifying statistical test for the multiple hypothesis tests was conducted. This test indicates that the overall variation in impacts across the implementation fidelity and duration subgroups is not statistically significant (F-statistic = 2.039; p-value = 0.086), suggesting that the statistical significance found on reading comprehension impacts for specific implementation fidelity or duration subgroups shown above should be interpreted cautiously.

Appendix Tables I.6 and I.7 correspond with the final panel in Table 5.6 and compare the impact estimates for the 15 schools with both (1) longer duration and (2) implementation fidelity that was classified as either well aligned or moderately aligned with the program model with the impact estimates for the 19 schools that had shorter program duration or implementation that was classified as poorly aligned with the program model. Appendix Table I.6 shows that the ERO programs produced positive and statistically significant impacts on reading comprehension in the schools that were both (1) well aligned or moderately aligned and (2) had longer duration. The difference between the impact on reading comprehension for these schools and the impact for the schools that faced more serious problems is a 0.16 effect size and is statistically significant. Appendix Table I.7 shows impacts on the amounts of school-related and non-school-related reading for programs with implementation that was well aligned or moderately aligned to the program model and had a longer duration. These impacts are not statistically significant.

The Enhanced Reading Opportunities Study

Appendix Table I.1

**Fixed-Effect Impact Estimates on Reading Comprehension,
by School**

Variable	Impact Estimate	Standard Error	95% Confidence Interval	
School 1 ^a	-7.1 *	2.37	-11.71	-2.40
School 2	-3.7	2.10	-7.83	0.39
School 3	-3.2	2.35	-7.84	1.40
School 4	-2.2	2.17	-6.41	2.09
School 5	-1.6	2.22	-5.93	2.78
School 6	-1.3	1.91	-5.04	2.47
School 7	-1.2	2.22	-5.60	3.10
School 8	-1.2	2.31	-5.72	3.35
School 9	-0.9	1.85	-4.56	2.72
School 10	-0.3	2.07	-4.40	3.73
School 11	-0.3	1.92	-4.08	3.46
School 12	0.2	2.48	-4.63	5.10
School 13	0.3	2.00	-3.66	4.18
School 14	0.4	2.51	-4.56	5.30
School 15	0.4	2.44	-4.42	5.17
School 16	0.6	2.53	-4.34	5.59
School 17	0.9	1.98	-3.00	4.77
School 18	0.9	2.46	-3.93	5.73
School 19	1.0	2.67	-4.25	6.22
School 20	1.2	2.08	-2.90	5.26
School 21	1.5	2.22	-2.81	5.90
School 22	1.6	2.75	-3.80	7.01
School 23	1.8	2.53	-3.12	6.81
School 24	2.1	1.97	-1.79	5.94
School 25	2.4	2.75	-3.00	7.80

(continued)

Appendix Table I.1 (continued)

Variable	Impact Estimate	Standard Error	95% Confidence Interval	
School 26	3.0	3.33	-3.56	9.48
School 27	3.3	2.06	-0.71	7.36
School 28	3.4	2.36	-1.23	8.05
School 29	3.5	1.88	-0.18	7.19
School 30	4.9	2.58	-0.18	9.93
School 31	5.0 *	2.36	0.42	9.66
School 32	5.1 *	2.20	0.81	9.43
School 33	5.7 *	1.90	2.00	9.45
School 34	5.9 *	2.24	1.49	10.28

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 2006 near the end of students' ninth-grade year.

The fixed-effect estimated impacts are the regression-adjusted impacts of the interaction between school and treatment 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.

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.

A composite F-test was used to assess whether the school-level impacts on reading comprehension test scores are statistically equivalent. The F-value is 1.63 and the p-value is 0.013, indicating that the school-to-school variation in impacts is unlikely to have occurred by chance.

^aThe schools are listed in ascending order by their impact estimate.

The Enhanced Reading Opportunities Study

Appendix Table I.2

Impacts on Reading Achievement,
Cohort 1 Follow-Up Respondent Sample,
by Program Implementation Fidelity

Outcome	Non-ERO ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Well-aligned implementation^a</u>					
Reading comprehension					
Average standard score	90.9	90.3	0.6	0.06	0.260
<i>Corresponding grade equivalent</i>	6.3	6.2			
<i>Corresponding percentile</i>	26	25			
Reading vocabulary					
Average standard score	93.8	94.3	-0.5	-0.05	0.404
<i>Corresponding grade equivalent</i>	7.8	7.8			
<i>Corresponding percentile</i>	33	34			
Sample size	633	455			
<u>Moderately aligned implementation</u>					
Reading comprehension					
Average standard score	90.0	87.7	2.3 *	0.22 *	0.005
<i>Corresponding grade equivalent</i>	6.1	5.5			
<i>Corresponding percentile</i>	25	19			
Reading vocabulary					
Average standard score	93.7	92.0	1.8 *	0.17 *	0.027
<i>Corresponding grade equivalent</i>	7.8	7.6			
<i>Corresponding percentile</i>	32	29			
Sample size	340	250			
<u>Poorly aligned implementation</u>					
Reading comprehension					
Average standard score	89.1	89.0	0.2	0.02	0.797
<i>Corresponding grade equivalent</i>	5.9	5.9			
<i>Corresponding percentile</i>	23	23			
Reading vocabulary					
Average standard score	92.7	92.4	0.3	0.03	0.655
<i>Corresponding grade equivalent</i>	7.7	7.6			
<i>Corresponding percentile</i>	30	30			
Sample size	435	300			

(continued)

Appendix Table I.2 (continued)

Difference in Impacts Between Subgroups	Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>Well-aligned minus poorly aligned</u>			
Reading comprehension standard score	0.4	0.04	0.636
Reading vocabulary standard score	-0.8	-0.08	0.385
<u>Moderately aligned minus poorly aligned</u>			
Reading comprehension standard score	2.2 *	0.21 *	0.050
Reading vocabulary standard score	1.5	0.14	0.177

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 2006 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

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 = 10.458; reading vocabulary = 10.505).

A two-tailed t-test was applied to the impact estimate and to the difference in impacts. 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.

^aThe fidelity of program implementation is measured on two dimensions: learning environment and comprehension instruction. On each dimension, schools were measured in terms of their depth of alignment to the program model. Schools that were well aligned to both dimensions are categorized as having "well-aligned implementation." Schools that were moderately aligned to at least one dimension and moderately or well aligned to the other dimension are categorized as being "moderately aligned." Schools that were poorly aligned to one or both dimensions are categorized as being "poorly aligned."

The Enhanced Reading Opportunities Study
Appendix Table I.3
Impacts on Reading Behaviors,
Cohort 1 Follow-Up Respondent Sample,
by Program Implementation Fidelity

Outcome	ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Well-aligned implementation^a</u>					
Amount of school-related reading (prior month occurrences)	40.2	38.4	1.8	0.04	0.466
Amount of non-school-related reading (prior month occurrences)	26.3	24.3	2.0	0.06	0.282
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	-0.02	0.778
Sample size	634	453			
<u>Moderately aligned implementation</u>					
Amount of school-related reading (prior month occurrences)	46.7	39.7	7.0	0.16	0.057
Amount of non-school-related reading (prior month occurrences)	28.4	24.2	4.2	0.13	0.120
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	-0.07	0.362
Sample size	339	251			
<u>Poorly aligned implementation</u>					
Amount of school-related reading (prior month occurrences)	47.8	53.5	-5.6	-0.13	0.115
Amount of non-school-related reading (prior month occurrences)	27.8	30.1	-2.3	-0.07	0.345
Use of reflective reading strategies (4-point scale)	2.7	2.6	0.0	0.06	0.433
Sample size	437	298			

(continued)

Appendix Table I.3 (continued)

Difference in Impacts Between Subgroups	Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>Well-aligned minus poorly aligned</u>			
Amount of school-related reading	7.5	0.17	0.087
Amount of non-school-related reading	4.2	0.13	0.160
Use of reflective reading strategies	0.0	-0.07	0.435
<u>Moderately aligned minus poorly aligned</u>			
Amount of school-related reading	12.6 *	0.29 *	0.014
Amount of non-school-related reading	6.5	0.20	0.073
Use of reflective reading strategies	-0.1	-0.13	0.229

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 43.867; non-school-related reading standard deviation = 31.834; use of reading strategies standard deviation = 0.670).

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

For each of the above measures, data are missing for no more than 5 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

^aThe fidelity of program implementation is measured on two dimensions: learning environment and comprehension instruction. On each dimension, schools were measured in terms of their depth of alignment to the program model. Schools that were well aligned to both dimensions are categorized as having “well-aligned implementation.” Schools that were moderately aligned to at least one dimension and moderately or well aligned to the other dimension are categorized as being “moderately aligned.” Schools that were poorly aligned to one or both dimensions are categorized as being “poorly aligned.”

The Enhanced Reading Opportunities Study

Appendix Table I.4

Impacts on Reading Achievement,
Cohort 1 Follow-Up Respondent Sample,
by Program Duration

Outcome	Non-ERO ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>More than 8.0 months^a</u>					
Reading comprehension					
Average standard score	90.8	89.2	1.7 *	0.16 *	0.039
<i>Corresponding grade equivalent</i>	6.3	5.9			
<i>Corresponding percentile</i>	26	23			
Reading vocabulary					
Average standard score	92.9	93.9	-1.0	-0.09	0.258
<i>Corresponding grade equivalent</i>	7.7	7.8			
<i>Corresponding percentile</i>	31	33			
Sample size	284	204			
<u>7.6 to 8.0 months</u>					
Reading comprehension					
Average standard score	89.3	88.3	1.0	0.10	0.081
<i>Corresponding grade equivalent</i>	6.0	5.7			
<i>Corresponding percentile</i>	23	21			
Reading vocabulary					
Average standard score	93.5	92.8	0.7	0.06	0.239
<i>Corresponding grade equivalent</i>	7.7	7.7			
<i>Corresponding percentile</i>	32	31			
Sample size	672	497			
<u>7.5 months or fewer</u>					
Reading comprehension					
Average standard score	91.0	90.8	0.2	0.02	0.712
<i>Corresponding grade equivalent</i>	6.3	6.3			
<i>Corresponding percentile</i>	26	26			
Reading vocabulary					
Average standard score	93.8	93.3	0.5	0.05	0.487
<i>Corresponding grade equivalent</i>	7.8	7.7			
<i>Corresponding percentile</i>	33	32			
Sample size	452	304			

(continued)

Appendix Table I.4 (continued)

Difference in Impacts Between Subgroups	Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>More than 8.0 months minus 7.5 months or fewer</u>			
Reading comprehension standard score	1.4	0.13	0.174
Reading vocabulary standard score	-1.5	-0.14	0.187
<u>7.6 to 8.0 months minus 7.5 months or fewer</u>			
Reading comprehension standard score	0.8	0.07	0.380
Reading vocabulary standard score	0.2	0.02	0.842

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 2006 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

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 = 10.458; reading vocabulary = 10.505).

A two-tailed t-test was applied to the impact estimate and to the difference in impacts. 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.

^aProgram duration refers to how long the ERO classes were in session during the school year.

The Enhanced Reading Opportunities Study
Appendix Table I.5
Impacts on Reading Behaviors,
Cohort 1 Follow-Up Respondent Sample,
by Program Duration

Outcome	Non-ERO ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>More than 8.0 months^a</u>					
Amount of school-related reading (prior month occurrences)	45.0	42.7	2.3	0.05	0.579
Amount of non-school-related reading (prior month occurrences)	27.7	26.1	1.6	0.05	0.593
Use of reflective reading strategies (4-point scale)	2.6	2.7	0.0	-0.06	0.482
Sample size	285	203			
<u>7.6 to 8.0 months</u>					
Amount of school-related reading (prior month occurrences)	47.3	47.6	-0.3	-0.01	0.922
Amount of non-school-related reading (prior month occurrences)	27.7	27.1	0.6	0.02	0.745
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	0.02	0.675
Sample size	673	494			
<u>7.5 months or fewer</u>					
Amount of school-related reading (prior month occurrences)	39.0	37.4	1.6	0.04	0.570
Amount of non-school-related reading (prior month occurrences)	26.4	24.2	2.2	0.07	0.339
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	-0.02	0.754
Sample size	452	305			

(continued)

Appendix Table I.5 (continued)

Difference in Impacts Between Subgroups	Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>More than 8.0 months minus 7.5 months or fewer</u>			
Amount of school-related reading	0.7	0.01	0.896
Amount of non-school-related reading	-0.6	-0.02	0.874
Use of reflective reading strategies	0.0	-0.04	0.729
<u>7.6 to 8.0 months minus 7.5 months or fewer</u>			
Amount of school-related reading	-1.9	-0.04	0.635
Amount of non-school-related reading	-1.6	-0.05	0.585
Use of reflective reading strategies	0.0	0.05	0.613

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 43.867; non-school-related reading standard deviation = 31.834; use of reading strategies standard deviation = 0.670).

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

For each of the above measures, data are missing for no more than 6 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

^aProgram duration refers to how long the ERO classes were in session during the school year.

The Enhanced Reading Opportunities Study

Appendix Table I.6

Impacts on Reading Achievement,
Cohort 1 Follow-Up Respondent Sample,
by First-year Implementation Issues

Outcome	Non-ERO ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Moderately or well-aligned implementation and longer duration^a</u>					
Reading comprehension					
Average standard score	90.7	89.0	1.8 *	0.17 *	0.002
<i>Corresponding grade equivalent</i>	6.2	5.9			
<i>Corresponding percentile</i>	26	23			
Reading vocabulary					
Average standard score	93.6	93.5	0.1	0.01	0.848
<i>Corresponding grade equivalent</i>	7.8	7.7			
<i>Corresponding percentile</i>	32	32			
Sample size	656	488			
<u>Poorly aligned implementation or shorter duration^b</u>					
Reading comprehension					
Average standard score	89.6	89.5	0.1	0.01	0.811
<i>Corresponding grade equivalent</i>	6.0	6.0			
<i>Corresponding percentile</i>	24	24			
Reading vocabulary					
Average standard score	93.3	92.9	0.4	0.04	0.412
<i>Corresponding grade equivalent</i>	7.7	7.7			
<i>Corresponding percentile</i>	32	31			
Sample size	752	517			
<u>Differences in impacts</u>					
Difference in Impacts Between Subgroups			Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
Reading comprehension standard score			1.6 *	0.16 *	0.035
Reading vocabulary standard score			-0.3	-0.03	0.667

(continued)

Appendix Table I.6 (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 2006 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

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 = 10.458; reading vocabulary = 10.505).

A two-tailed t-test was applied to the impact estimate and to the difference in impacts. 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.

^aThe ERO programs in these schools were deemed to have reached an implementation level that was moderately or well aligned to both the classroom learning environment and comprehension instruction dimensions of the program model, and they were in operation for more than 7.5 months.

^bThe implementation fidelity of the ERO programs in these schools was deemed to be poorly aligned to the classroom learning environment and/or comprehension instruction dimensions of the program model, and/or they were in operation for 7.5 months or less.

The Enhanced Reading Opportunities Study
Appendix Table I.7
Impacts on Reading Behaviors,
Cohort 1 Follow-Up Respondent Sample,
by First-year Implementation Issues

Outcome	Non-ERO ERO	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Moderately or well-aligned implementation and longer duration^a</u>					
Amount of school-related reading (prior month occurrences)	45.4	40.5	4.9	0.11	0.065
Amount of non-school-related reading (prior month occurrences)	28.1	24.8	3.3	0.10	0.075
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	0.01	0.887
Sample size	656	486			
<u>Poorly aligned implementation or shorter duration^b</u>					
Amount of school-related reading (prior month occurrences)	43.2	46.0	-2.9	-0.07	0.250
Amount of non-school-related reading (prior month occurrences)	26.5	27.1	-0.6	-0.02	0.744
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	-0.02	0.695
Sample size	754	516			
<hr/>					
Difference in Impacts Between Subgroups			Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>Differences in impacts</u>					
Amount of school-related reading			7.7 *	0.18 *	0.033
Amount of non-school-related reading			3.9	0.12	0.129
Use of reflective reading strategies			0.0	0.03	0.709

(continued)

Appendix Table I.7 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at 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 ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 43.867; non-school-related reading standard deviation = 31.834; use of reading strategies standard deviation = 0.670).

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

For each of the above measures, data are missing for no more than 4.7 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

^aThe ERO programs in these schools were deemed to have reached an implementation level that was moderately or well aligned to both the classroom learning environment and comprehension instruction dimensions of the program model, and they were in operation for more than 7.5 months.

^bThe implementation fidelity of the ERO programs in these schools was deemed to be poorly aligned to the classroom learning environment and/or comprehension instruction dimensions of the program model, and/or they were in operation for 7.5 months or less.

References

- Abrams, Andrew, and Diana Oxley (eds.). 2006. *Critical Issues in Development and Implementation of High School Small Learning Communities*. Washington, DC: U.S. Department of Education.
- Alliance for Excellent Education. 2004. *How to Know a Good Adolescent Literacy Program When You See One: Quality Criteria to Consider*. Washington, DC: Alliance for Excellent Education. Web site: http://www.all4ed.org/adolescent_literacy/issue_briefs.html.
- Alvermann, Donna. 2002. "Effective Literacy Instruction for Adolescents." *Journal of Literacy Research* 34: 189-208.
- 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.
- American Institutes for Research. 2004. "Request for Proposals from Vendors of High Quality Supplemental Literacy Programs for Striving Readers in Ninth Grade." Available from the author on request at ERO@air.org.
- Balfanz, Robert, and Nettie Legters. 2004. *Locating the Dropout Crisis*. Baltimore: Center for Social Organization of Schools. Web site: http://www.csos.jhu.edu/tdhs/rsch/Locating_Dropouts.pdf.
- Beck, Isabel, Margaret McKeown, and Linda Kucan. 2002. *Bringing Words to Life: Robust Vocabulary Instruction*. New York: Guilford.
- Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling for the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B (Methodological)* 57, 1: 289-300.
- Biancarosa, Gina, and Catherine E. Snow. 2004. *Reading Next — A Vision for Action and Research in Middle and High School Literacy: A Report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education. Web site: <http://www.all4ed.org/publications/ReadingNext/ReadingNext.pdf>.
- Bloom, Harold, Carolyn Hill, Alison Rebeck Black, and Mark Lipsey. 2006. "Effect Sizes in Education Research: What They Are, What They Mean, and Why They Are Important." Presentation for the Institute of Education Sciences 2006 Research Conference. New York: MDRC.
- Bloome, David. 2001. "Boundaries on the Construction of Literacy in Secondary Classrooms: Envisioning Reading and Writing in a Democratic and Just Society." Pages 287-304 in Elizabeth Birr Moje and David G. O'Brien (eds.), *Constructions of Literacy: Studies of Teaching and Learning In and Out of Secondary Schools*. Mahwah, NJ: Erlbaum.

- Carnevale, Anthony P. 2001. *Help Wanted...College Required*. Washington, DC: Educational Testing Service, Office for Public Leadership.
- Cronbach, Lee J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16: 297-334.
- Curtis, Mary. E., and Mary B. Chmelka. 1994. "Modifying the Laubach Way To Reading Program for Use with Adolescents with LDs." *Learning Disabilities: Research and Practice* 9: 38-43.
- Darling-Hammond, Linda, Jacqueline Aness, and Susanna Ort. 2002. "Reinventing High School: Outcomes of the Coalition Campus Schools Project." *American Education Research Journal* 39: 639-673.
- Dillon, Deborah R., David G. O'Brien, and Mark J. Volkmann. 2001. "Reading, Writing, and Talking to Get Work Done in Biology." Pages 51-76 in Elizabeth Birr Moje and David G. O'Brien (eds.), *Constructions of Literacy: Studies of Teaching and Learning In and Out of Secondary Schools*. Mahwah, NJ: Erlbaum.
- Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. *Using Randomization in Development Economics: A Toolkit*. London: Centre for Economic Policy Research. Web site: <http://www.cepr.org/pubs/dps/DP6059.asp>.
- Fine, Michelle. 1988. "Deinstitutionalizing Educational Inequity." Pages 88-119 in Council of Chief State School Officers (eds.), *School Success for Students at Risk*. New York: Harcourt.
- Guthrie, John T. 2002. *Proceedings of Adolescent Literacy — Research Informing Practice: A Series of Workshops*. Washington, DC: National Institute for Literacy. Web site: <http://www.nifl.gov/partnershipforreading/publications/adolescent.html>.
- Guthrie, John T., and Donna Alvermann (eds.). 1999. *Engaged Reading: Processes, Practices and Policy Implications*. New York: Teachers College Press.
- Harvey, James, and Naomi Housman. 2004. *Crisis or Possibility? Conversations About the American High School*. Washington, DC: National High School Alliance.
- Hedges, Larry V., and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis*. San Diego, CA: Academic Press.
- Kamil, Michael L. 2003. *Adolescents and Literacy: Reading for the 21st Century*. Washington, DC: Alliance for Excellent Education. Web site: <http://web.all4ed.org/publications/AdolescentsAndLiteracy.pdf>.
- Kemple, James, James P. Connell, Nettie Legters, and Jacquelynne Eccles. 2006. "Making the Move: How Freshman Academies and Thematic Small Learning Communities Can Support Successful Transitions to and Through High School." Pages 235-285 in Andrew Abrams and Diana Oxley (eds.), *Critical Issues in Development and Implementation of High School Small Learning Communities*. Washington, DC: U.S. Department of Education.

- Kemple, James, and Corinne Herlihy. 2004. *The Talent Development High School Model: Context, Components, and Initial Impacts on Ninth-Grade Students' Engagement and Performance*. New York: MDRC.
- Kling, Jeffrey R., and Jeffrey B. Liebman, 2004. "Experimental Analysis of Neighborhood Effects on Youth." KSG Working Paper RWP04-034. Cambridge, MA: Harvard University, John F. Kennedy School of Government.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75, 1: 83-119.
- Lee, Valerie E., Anthony S. Bryk, and Julia B. Smith. 1993. "The Organization of Effective Secondary Schools." Pages 171-267 in Linda Darling-Hammond (ed.), *Review of Research in Education: Vol. 19*. Washington, DC: American Educational Research Association.
- Legters, Nettie, and Kerri Kerr. 2001. "Easing the Transition to High School: An Investigation of Reform Practices to Promote Ninth Grade Success." Center for Social Organization of Schools, Johns Hopkins University. Prepared for a forum convened by the Civil Rights Project at Harvard University's Graduate School of Education and by Achieve, Inc.
- Lipsey, Mark. 1990. *Design Sensitivity: Statistical Power for Experimental Research*. Newbury Park, CA: Sage.
- Logan, Brent R., and Ajit C. Tamhane. 2003. "On O'Brien's OLS and GLS Tests for Multiple Endpoints." Department of Industrial Engineering and Management Sciences Working Paper 03-0004. Evanston, IL: Northwestern University.
- Lutkus, Anthony D., Bobby D. Rampey, and Patricia L. Donahue. 2006. *The Nation's Report Card: Trial Urban District Assessment Reading 2005*. NCES 2006-455r. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- National Association of Secondary School Principals (NAESP). 1996. *Breaking Ranks: Changing an American Institution*. Alexandria, VA: National Association of Secondary School Principals.
- National Center for Education Statistics. 1990. *National Education Longitudinal Study of 1998: A profile of the American Eighth Grader: NELS: -88 Student Descriptive Summary*. Washington, DC: U.S. Department of Education.
Web site: <http://nces.ed.gov/pubs90/90458.pdf>.
- National Reading Panel. 2000. *Report of the National Reading Panel: Teaching Children to Read*. Washington, DC: National Institute of Child Health and Human Development, National Institutes of Health, U.S. Department of Health and Human Services.
Web site: <http://www.nichd.nih.gov/publications/nrppubskey.cfm>.
- O'Brien, Peter C. 1984. "Procedures for Comparing Samples with Multiple Endpoints." *Biometrics* 40: 1079-1087.
- O'Brien, David G., Elizabeth Birr Moje, and Roger A. Stewart. 2000. "Exploring the Context of Secondary Literacy: Literacy in People's Everyday School Lives." Pages 27-48 in Elizabeth

- Birr Moje and David G. O'Brien (eds.), *Constructions of Literacy: Studies of Teaching and Learning In and Out of Secondary Schools*. Mahwah, NJ: Erlbaum.
- Peterson, Cynthia L., David C. Caverly, Sheila A. Nicholson, Sharon O'Neal, and Susen Cusenbary. 2000. *Building Reading Proficiency at the Secondary Level: A Guide to Resources*. Austin, TX: Southwest Educational Development Laboratory.
- RAND Reading Study Group. 2002. *Reading for Understanding: Toward a Research and Development Program in Reading Comprehension*. Santa Monica, CA: RAND Reading Study Group.
- Roderick, Melissa. 1993. *The Path to Dropping Out: Evidence for Intervention*. Westport, CT: Auburn House.
- Roe, Betty D., Barbara D. Stoodt, and Paul C. Burns. 1998. *Secondary School Literacy Instruction: The Content Areas* (6th ed.). Boston: Houghton Mifflin.
- Quint, Janet. 2006. *Meeting Five Critical Challenges of High School Reform: Lessons from Research on Three Reform Models*. New York: MDRC.
- Quint, Janet, Cynthia Miller, Jennifer Pastor, and Rachel Cytron. 1999. *Project Transition: Testing an Intervention to Help High School Freshmen Succeed*. New York: MDRC.
- Schoenbach, Ruth, Cynthia Greenleaf, Christine Cziko, and Lori Hurwitz. 1999. *Reading for Understanding: A Guide to Improving Reading in Middle and High School Classrooms*. San Francisco: Jossey-Bass.
- Schumaker, Jean B., and Donald D. Deshler. 2003. "Designs for Applied Educational Research." Pages 283-500 in H. Lee Swanson, Karen R. Harris, and Steve Graham (eds.), *Handbook of Learning Disabilities*. New York: Guilford.
- Schumaker, Jean B., and Donald D. Deshler. 2004. "Teaching Adolescents to be Strategic Learners." In Donald D. Deshler and Jean B. Schumaker (eds.), *High School Students with Disabilities: Strategies for Accessing the Curriculum*. New York: Corwin.
- Shanahan, Timothy. 2004. "Improving Reading Achievement in Secondary Schools: Structures and Reforms." Pages 43-55 in Donna Alvermann and Dorothy Strickland (eds.), *Bridging the Literacy Achievement Gap: Grades 4-12*. New York: Teachers College Press.
- Sizer, Theodore. 1984. *Horace's Compromise: The Dilemma of the American High School*. Boston, MA: Houghton Mifflin.
- Snow, Catherine E., and Gina Biancarosa. 2003. *Adolescent Literacy and the Achievement Gap: What Do We Know and Where Do We Go from Here?* New York: Carnegie Corporation.
- University of Kansas Center for Research on Learning (KU-CRL). 2004. "SIM Adolescent Reading Project: Technical Proposal." Unpublished proposal. Lawrence, KS: University of Kansas Center for Research on Learning.

- U.S. Department of Education. 2005. *Smaller Learning Communities: Special Competition for Supplemental Reading Program Research Evaluation*. Washington DC: U.S. Department of Education.
- WestEd. 2004. "High Quality Supplemental Literacy Programs for Striving Readers in Ninth Grade." Unpublished proposal. San Francisco: WestEd.
- Wigfield, Allen. 2004. "Motivation for Reading During the Early Adolescent and Adolescent Years." Pages 56-69 in Dorothy Strickland and Donna Alvermann (eds.), *Bridging the Literacy Achievement Gap, Grades 4-12*. New York: Teachers College Press.