NCEE 2010-4021

U. S. DEPARTMENT OF EDUCATION

The Enhanced Reading Opportunities Study Final Report

The Impact of Supplemental Literacy Courses for Struggling Ninth-Grade Readers

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

The Enhanced Reading Opportunities Study Final Report

The Impact of Supplemental Literacy Courses for Struggling Ninth-Grade Readers

July 2010

Marie-Andrée Somers
William Corrin
Susan Sepanik
MDRC

Terry Salinger
Jesse Levin
Courtney Zmach
American Institutes for Research

With

Edmond Wong
MDRC

Paul Strasberg, Project Officer
Marsha Silverberg, Project Officer
Institute of Education Sciences

NCEE 2010-4021
U.S. Department of Education

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Evaluation and Regional Assistance
John Q. Easton
Acting Commissioner

July 2010

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: Somers, M.-A., Corrin, W., Sepanik, S., Salinger T., Levin, J., and Zmach, C. (2010). *The Enhanced Reading Opportunities Study Final Report: The Impact of Supplemental Literacy Courses for Struggling Ninth-Grade Readers* (NCEE 2010-4021). 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 22207, Alexandria, VA 22304.
- 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 877-576-7734.
- Fax your request to 703-605-6794 or order online at www.edpubs.org.

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

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 202-260-9895 or 202-205-8113.

Contents

List of Exhibits v
Acknowledgments xiii
Disclosure of Potential Conflicts of Interest xv
Executive Summary ES-1

Chapter

1 **Introduction** 1

Overview of the ERO study 3
Overview of the Supplemental Literacy Interventions 3
Implementation of the Programs and Teacher Training 4
A Rigorous Impact Evaluation 6
Overview of This Report 9

2 **Study Design: Sample, Measures, and Analysis** 11

Schools in the Study 12
Random Assignment of Schools 13
Characteristics of Schools Selected for the ERO Project 14
Students in the Study 16
Random Assignment of Students 17
Characteristics of Students in the Study 17
Data Sources and Measures 18
Student Outcomes and Baseline Characteristics 18
Implementation Fidelity and Service Contrast 28
Analysis Samples in the Study 29
Availability of Data 29
Students in the Analysis Samples 32
Analytic Methods and Procedures 37
Statistical Model for Estimating Impacts 37
Sample Sizes and Statistical Power 41
Statistical Significance and Multiple Hypothesis Testing 46

3 Program Implementation and Impacts on Reading Achievement and Reading Behaviors 49

Program Implementation 49
Implementation Fidelity: Did the ERO Programs as Implemented Match the Intent of Their Design? 50
Program Duration and Student Attendance: How Much ERO Instruction Did ERO Students Receive? 57
Student Course Schedules and Literacy Support Services: How Different Were the ERO Courses from "Business as Usual"? 59
Program Impacts on Reading Achievement and Reading Behaviors 63
Impacts on Student Reading Achievement: Did the ERO Program Improve Students' Reading Skills? 63

Impacts on Student Reading Behaviors: Did the ERO Program Change How and How Much Students Read? 68

4 Impacts on High School Outcomes 73

Overall Impacts 74
Course Performance 75
State Test Scores 84
Student Behaviors 86
Impacts, by Student Subgroup 88
Impacts, by Program 88
Course Performance 89
State Test Scores 93
Student Behaviors 96

5 Exploring Hypotheses About Program Effects and Their Persistence 99

The Relationship Between Implementation Characteristics and Program Impacts 99
Variation in Impacts Across Schools 99
Association Between Impacts and Implementation Characteristics 105

6 Program Costs and Poststudy Implementation 109

The Cost of Implementing the ERO Programs 110
Poststudy Continuation and Modification of the ERO Programs in the Study Schools 114
Estimated Program Costs Under Nonstudy Conditions 118

References 123

Appendixes

A The ERO Programs and the ERO Teachers A-1
B ERO Student Survey Measures B-1
C ERO Implementation Fidelity C-1
D State Tests Included in the ERO Study D-1
E Response Analysis and Baseline Comparison Tables E-1
F Technical Notes for Impact Findings F-1
G Statistical Power and Minimum Detectable Effect Size G-1
H Supplementary Impact Findings H-1
I Baseline and Impact Findings, by Cohort I-1
J The Association Between Reading Outcomes and Academic Performance in High School J-1
K Variation in Impacts Across Sites and Cohorts K-1
L Program Costs L-1
M Poststudy Adolescent Literacy Programming in the ERO Schools: Methodology and Additional Findings M-1

List of Exhibits

Table

ES.1	Impacts on State Test Scores (Standardized), School Records Sample	ES-15
ES.2	Impact on Grade Point Average (GPA) and Credit Accumulation in Core Subject Areas, (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample	ES-17
2.1	Characteristics of ERO Schools and Average Schools in the U.S. (2004-2005)	15
2.2	Data Collection Years, by Cohort	20
2.3	Student Outcomes in the ERO Study	21
2.4	Availability of Data for the Full Study Sample, by Data Source and Type of Outcome	30
2.5	Characteristics of Students in the Analysis Samples Relative to Students in the Full Study Sample	35
2.6	Minimum Detectable Effect (MDE) and Effect Size (MDES) for Impacts on Reading Achievement and Reading Behaviors, Follow-up Respondent Samples (GRADE and Student Survey)	43
2.7	Minimum Detectable Effect (MDE) and Effect Size (MDES) for Impacts on Course Performance and School Behaviors Outcomes, School Records Sample	44
2.8	Minimum Detectable Effect Size (MDES) for Impacts on State Test Performance, School Records Sample	48
3.1	Number of ERO Classrooms with Well-, Moderately, or Poorly Aligned Implementation to Program Models on Each Implementation Dimension, by ERO Program: Year 2, Spring	52
3.2	Attendance in ERO Classes, GRADE Respondent Sample in the ERO Group	58
3.3	Impacts on Reading Achievement, GRADE Respondent Sample, Two Cohorts Combined	65
3.4	Impacts on Reading Achievement, GRADE Respondent Sample, Two Cohorts Combined, by Program	66
3.5	Impacts on Reading Behaviors, GRADE Respondent Sample, Two Cohorts Combined	70
3.6	Impacts on Reading Behaviors, GRADE Respondent Sample, Two Cohorts Combined, by Program	71

Table

4.1	Impacts on Grade Point Average (GPA), School Records Sample	76
4.2	Impacts on Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample	80
4.3	Impacts on State Test Scores (Standardized), School Records Sample	85
4.4	Impacts on Student Behaviors, School Records Sample	87
4.5	Impacts on Grade Point Average (GPA), School Records Sample, by Program	90
4.6	Impacts on Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample, by Program	92
4.7	Impacts on State Test Scores, School Records Sample, by Program	94
4.8	Impacts on Student Behaviors, School Records Sample, by Program	97
5.1	Association Between Program Implementation Conditions and Impacts on Grade Point Average (GPA) in Core Subject Areas (Standardized Regression Coefficients)	107
6.1	Cost per Student of the ERO Programs, by Implementation Year (2008-2009 Dollars)	112
6.2	Projected Cost per Student of the ERO Programs in a Hypothetical Nonstudy Setting, by Teacher Qualifications and Total Student Enrollment (2008-2009 Dollars)	121
A.1	Key Components of the ERO Programs	A-5
A.2	Comparison of ERO and Non-ERO Student Schedules	A-10
A.3a	Background Characteristics of ERO Teachers, First Year of Implementation	A-12
A.3b	Background Characteristics of ERO Teachers, Second Year of Implementation	A-13
A.4	Training and Technical Assistance Provided During Each School Year, by ERO Program	A-16
B.1	Intensity Values for Supplementary Literacy Support Measures	B-5
C.1	Number of ERO Classrooms with Well-, Moderately, or Poorly Aligned Implementation to Program Models on Each Implementation Dimension, by ERO Program: Year 1, Spring Site Visit	C-13
C.2	Number of ERO Classrooms with Well-, Moderately, or Poorly Aligned Implementation to Program Models on Each Implementation Dimension, by ERO Program: Year 2, Fall Site Visit	C-14

Table

C.3	Number of ERO Classrooms with Well-, Moderately, or Poorly Aligned Implementation to Program Models on Each Implementation Dimension, by ERO Program: Year 2, Fall and Spring Site Visits	C-15
C.4	Average Implementation Composite Scores, by Experience with the ERO Program: Year 2, Fall	C-16
C.5	Average Implementation Composite Scores, by Experience with the ERO Program, Year 2, Spring	C-17
D.1	Description of State Tests, by District	D-4
E.1	Characteristics of the Full Study Sample	E-4
E.2	Characteristics of the Full Study Sample, Reading Apprenticeship Schools	E-6
E.3	Characteristics of the Full Study Sample, Xtreme Reading Schools	E-8
E.4	Availability of Data for the Full Study Sample, by Data Source and Type of Outcome, Reading Apprenticeship Schools	E-11
E.5	Availability of Data for the Full Study Sample, by Data Source and Type of Outcome, Xtreme Reading Schools	E-12
E.6a	Characteristics of Students: Differences Between Students With and Without School Records in the Program Year	E-15
E.6b	Characteristics of Students: Differences Between Students With and Without School Records in the Follow-up Year	E-17
E.7a	Characteristics of Students: Differences Between Students With and Without School Records in the Program Year, Reading Apprenticeship Schools	E-19
E.7b	Characteristics of Students: Differences Between Students With and Without School Records in the Follow-Up Year, Reading Apprenticeship Schools	E-21
E.8a	Characteristics of Students: Differences Between Students With and Without School Records in the Program Year, Xtreme Reading Schools	E-23
E.8b	Characteristics of Students: Differences Between Students With and Without School Records in the Follow-up Year, Xtreme Reading Schools	E-25
E.9	Characteristics of Students with State Test Data Relative to Students in the Full Study Sample	E-27
E.10	Characteristics of Students with State Test Data Relative to Students in the Full Study Sample, Reading Apprenticeship Schools	E-29

Table

E.11	Characteristics of Students with State Test Data Relative to Students in the Full Study Sample, Xtreme Reading Schools	E-31
E.12a	Characteristics of Students with School Records in the Program Year	E-34
E.12b	Characteristics of Students with School Records in the Follow-Up Year	E-36
E.13a	Characteristics of Students with School Records in the Program Year, Reading Apprenticeship Schools	E-38
E.13b	Characteristics of Students with School Records in the Follow-Up Year, Reading Apprenticeship Schools	E-40
E.14a	Characteristics of Students with School Records in the Program Year, Xtreme Reading Schools	E-42
E.14b	Characteristics of Students with School Records in the Follow-Up Year, Xtreme Reading Schools	E-44
E.15	Tests for Systematic Differences in Baseline Characteristics Between the ERO and non-ERO Group, Students in the School Records Sample with State Test Scores	E-46
F.1	Impacts with Confidence Intervals on Reading Achievement, GRADE Respondent Sample	F-5
F.2	Impacts with Confidence Intervals on Grade Point Average (GPA) and Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample	F-6
F.3	Impacts with Confidence Intervals on Grade Point Average (GPA) and Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample, by Program	F-7
F.4	Impacts on Grade Point Average (GPA) and Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), Not Adjusted for Student Baseline Characteristics, School Records Sample	F-9
F.5	Impacts on Grade Point Average (GPA) and Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), Not Adjusted for Student Baseline Characteristics, School Records Sample, by Program	F-10
F.6	P-Values for Estimated Impacts on Primary Outcomes, Adjusted for Multiple Hypothesis Testing	F-13
F.7	Standard Deviations for the Non-ERO group, ERO Group, and Pooled Groups, by Outcome and Year	F-20

Table

G.1	School-Level and District-Level Intraclass Correlations for Student Outcomes	G-8
G.2	Variance Explained by Students' Background Characteristics and Prior Achievement	G-10
H.1	Impacts on Total Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample, by Program	H-4
H.2	Credits Attempted in Core Subject Areas, School Records Sample	H-6
H.3	Credits Attempted in Core Subject Areas, School Records Sample, by Program	H-7
H.4	Impacts on Meeting Proficiency Standards on State Tests, School Records Sample, Program Year	H-9
H.5	Impacts on Grade Point Average (GPA) During the Program Year: Follow-Up Year School Records Sample	H-11
H.6	Impacts on Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), in the Program Year: Follow-Up Year School Records Sample	H-13
H.7	Impacts on Student Behaviors During the Program Year, Follow-up Year School Records Sample	H-15
H.8	Impacts on Grade Point Average (GPA) and Credit Accumulation in Core Subject Areas, (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample, by Baseline Reading Comprehension Performance	H-17
H.9	Impacts on Grade Point Average (GPA) and Credit Accumulation in Core Subject Areas, (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample, by Whether Students Were Overage for Grade	H-19
H.10	Impacts on Grade Point Average (GPA) and Credit Accumulation in Core Subject Areas, (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample, by Language Spoken at Home	H-21
I.1	Characteristics of Students with School Records in the Program Year, Cohort 1	I-5
I.2	Characteristics of Students with School Records in the Follow-Up Year, Cohort 1	I-7
I.3	Characteristics of Students with School Records in the Second Follow-Up Year, Cohort 1	I-9
I.4	Characteristics of Students with School Records in the Program Year, Cohort 2	I-11
I.5	Characteristics of Students with School Records in the Follow-Up Year, Cohort 2	I-13

Table

I.6	Impacts on Grade Point Average (GPA) and Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), for Cohort 1, School Records Sample	I-15
I.7	Impacts on Grade Point Average (GPA) and Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), for Cohort 2, School Records Sample	I-16
J.1	Association Between Reading Outcomes and Grade Point Average in Core Subject Areas (Standardized Regression Coefficients)	J-6
J.2	Association Between Reading Outcomes and Credit Accumulation in Core Subject Areas (Standardized Regression Coefficients)	J-8
J.3	Association Between Reading Outcomes and Performance on State Tests (Standardized Regression Coefficients)	J-10
J.4	Correlations Between Reading Outcomes and Grade Point Average in Core Subject Areas	J-13
J.5	Correlations Between Reading Outcomes and Credit Accumulation in Core Subject Areas	J-15
J.6	Correlations Between Reading Outcomes and Performance on State Tests	J-17
K.1	Association Between Program Implementation Conditions and Impacts on Credit Accumulation in Core Subject Areas (Standardized Regression Coefficients)	K-12
L.1	Nonsalary Costs by Implementation Year and Cost Category (in Nominal Dollars)	L-6
L.2	Average Cost per Student, Number of Students Served, and Teacher Qualifications in Highest-, Middle-, and Lowest-Spenders Program Demonstration Schools (Figures Calculated Over Both Implementation Years)	L-8
L.3	Teacher Salary, by Years of Experience and Highest Degree Earned in 2008-2009 Dollars	L-9
M.1	Summary of ERO Continuation Plans, by District	M-6

Figure

ES.1	Impacts on Reading Achievement, GRADE Respondent Sample	ES-10
ES.2	Impacts on Grade Point Average (GPA), School Records Sample	ES-12
ES.3	Impacts on Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample	ES-13

Figure

1.1	The ERO Programs' Theory of Change	5
1.2	Study Timeline: Implementation Years	7
3.1	Learning Environment Composite Scores, by ERO Program	53
3.2	Comprehension Instruction Composite Scores, by ERO Program	54
3.3	Composite Fidelity Scores, by Site Visit	56
3.4	Participation in Supplemental Literacy Support Activities	61
4.1	Impacts on Grade Point Average (GPA), School Records Sample	77
4.2	Impacts on Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample	83
5.1	Fixed-Effect Impact Estimates on Grade Point Average (GPA) in Core Subject Areas in the Program Year, by School, School Records Sample	101
5.2	Fixed-Effect Impact Estimates on Grade Point Average (GPA) in Core Subject Areas in the Follow-Up Year, by School, School Records Sample	103
6.1	Continuation of the ERO Programs in the Study Schools After the Study-Required Implementation Period	115
6.2	Most Common Modifications to the ERO Programs in Schools That Continued to Use the Programs After the Study-Required Implementation Period	116
E.1	Creation of the Primary Analysis Sample	E-14
F.1	Density of GRADE Reading Comprehension Scores in the Program Year, by Program Group	F-14
F.2a	Density of Grade Point Average (GPA) in Core Subject Areas in the Program Year, by Program Group	F-15
F.2b	Density of Grade Point Average (GPA) in Core Subject Areas in the Follow-Up Year, by Program Group	F-16
F.3a	Density of Core Credit Accumulation in the Program Year, by Program Group	F-17
F.3b	Density of Core Credit Accumulation in the Follow-Up Year, by Program Group	F-18
K.1	Fixed-Effect Impact Estimates on Credit Accumulation in Core Subject Areas in the Program Year, by School, School Records Sample	K-5
K.2	Fixed-Effect Impact Estimates on Credit Accumulation in Core Subject Areas in the Follow-Up Year, by School, School Records Sample	K-7

Box

ES.1 Overview of the Study ES-4

2.1 Description of the Calculation and Presentation of Outcome Levels 40

Acknowledgments

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

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, Suzannah Herrmann and Kathryn Drummond conducted site visits and phone interviews for the implementation study. Christopher Doyle and Andrea Olinger coordinated data management and conducted phone interviews. Nancy Lang conducted phone interviews and site visits, processed data, and ensured the thoroughness of the implementation fidelity ratings. Eva Lyman-Munt conducted phone interviews with district and school staff for this report. Midori Rome conducted phone interviews and assisted with the analysis of interview data. Seth Brown conducted phone interviews, participated in collecting and analyzing the data for the cost study, and participated in writing up the results.

At MDRC, James Kemple served as the project director during the first few years of the study, and, in that role, he provided the strong leadership needed to get the project off the ground and a sturdy foundation for all aspects of the study, from school and student recruitment to the research design and analysis. Elizabeth Nelson was involved in all phases of program implementation, including site relations, student recruitment, and conducting implementation research. Alma Moedano, Alixandra Barasch, and Asa Wilks assisted with data collection and provided programming and analysis support. Corinne Herlihy and Kristin Porter served as school district coordinators. Daniel Fallon oversaw the ordering of the literacy assessment. Shirley James and her staff entered data. Gordon Berlin, Alison Black, Howard Bloom, Fred Doolittle, Corinne Herlihy, John Hutchins, Robert Ivry, Janet Quint, Michael Weiss, and Pei Zhu provided substantive expertise through their thoughtful comments on, and reviews of, this report. Alma Moedano, Julia Gomez, Asya Magazinnik, and Mario Flecha assisted with report production. Margaret Bald, John Hutchins, Alice Tufel, and Robert Weber edited the report, and Stephanie Cowell and Inna Kruglaya prepared it for publication.

The Authors

Disclosure of Potential Conflicts of Interest[1]

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 in an early panel meeting (2005) on the study design and a review meeting (2009) about the findings presented in this report. During the course of the study, a commercial literacy intervention he had developed was on the market that was targeted to striving middle-school readers. This intervention 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, but it is no longer actively marketed. Dr. Shanahan had no role in the selection of the study programs or in the analysis of evaluation data.

[1]Contractors 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

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.[1] Of these students, nearly half do not exhibit even partial mastery of the knowledge and skills that are fundamental to proficient work at grade level.[2] These limitations in literacy skills are a major source of course failure, high school dropout, and poor performance in postsecondary education.[3] 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.[4]

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 whose reading skills are at least two years below grade level.[5] As part of this demonstration, 34 high schools from 10 school districts implemented one of two reading interventions: Reading Apprenticeship Academic Literacy (RAAL), designed by WestEd, and Xtreme Reading, designed by the University of Kansas Center for Research on Learning. These programs were implemented in the study schools for two school years. The U.S. Department of Education's (ED) Office of Elementary and Secondary Education (OESE)[6] funded the implementation of these programs, and its Institute of Education Sciences (IES) was responsible for oversight of the evaluation. MDRC — a nonprofit, nonpartisan education and social policy research organization — conducted the evaluation in partnership with the American Institutes for Research (AIR) and Survey Research Management (SRM).

The goal of the reading interventions — which consist of a year-long course that replaces a ninth-grade elective class — is to help striving adolescent readers develop the strategies and routines used by proficient readers, thereby improving their reading skills and ultimately,

---

[1]The NAEP is a nationally representative assessment of student achievement overseen by the National Center for Education Statistics (http://nces.ed.gov/nationsreportcard/).

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

[3]Carnevale (2001); Kamil (2003); Snow and Biancarosa (2003).

[4]Biancarosa and Snow (2004).

[5]The ERO study is known more formally as "An Evaluation of the Impact of Supplemental Literacy Interventions in Freshman Academies."

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

their academic performance in high school. The first two reports for the study evaluated the programs' impact on the two most proximal outcomes targeted by the interventions — students' reading skills and their reading behaviors at the end of ninth grade.[7] This report — which is the final of three reports for this evaluation — examines the impact of the ERO programs on the more general outcomes that the programs hope to affect — students' academic performance in high school (grade point average [GPA], credit accumulation, and state test scores) as well as students' behavioral outcomes (attendance and disciplinary infractions). These academic and behavioral outcomes are examined during the year in which they were enrolled in the ERO programs (ninth grade), as well as the following school year (tenth grade for most students).

Overall, the findings from these reports show that over the course of ninth grade, the ERO programs improved students' reading comprehension skills and helped them perform better academically in their high school course work. However, these benefits did not persist in the following school year, when students were no longer receiving the supports provided by the ERO programs. The key findings from the study are the following:[8]

- **The ERO programs improved students' reading comprehension skills over the course of ninth grade.** Across both cohorts of participating ninth-grade students, the ERO programs improved students' reading comprehension scores by an effect size of 0.09,[9] corresponding to an improvement from the twenty-third percentile to the twenty-fifth percentile nationally. However, 77 percent of students assigned to the ERO classes were still reading at two or more years below grade level at the end of ninth grade.

- **During the ninth grade, the ERO programs also had a positive impact on students' academic performance in core subject areas.** Students' GPA in core subject areas[10] was 0.06 point higher (out of a maximum of 4 points) as a result of being assigned to the ERO program (effect size = 0.07). The programs also helped students earn 0.6 percentage point more of the core credits that they need to graduate (effect size = 0.06). In the subset of high schools located in states where standardized tests are administered in ninth grade, students also scored higher on their English lan-

[7]The first report presented implementation and impact findings for the first year of program operations (Kemple et al., 2008), while the second report focused on impacts in the second year of implementation (Corrin et al., 2008). Chapter 3 of this report provides a review of the implementation and impact findings from these two reports and also presents the average impact of the reading programs across both cohorts.

[8]The statistical significance of all impact estimates in this report is evaluated at the 5 percent level.

[9]In this report, effect sizes are expressed as a proportion of the overall variability (standard deviation) in the outcome measure among students who were *not* assigned to the reading program (non-ERO group).

[10]The four core subject areas are English language arts, social studies, science, and mathematics.

guage arts and mathematics tests as a result of having been assigned to the ERO program; the estimated effect size of these impacts are 0.11 and 0.07, respectively.

- **However, in the school year following students' participation in the ERO programs, the programs no longer had an impact on academic performance.** Estimated impacts on students' GPA in core subject areas, credit accumulation, and standardized state test scores are not statistically significant in the school year following program participation (tenth grade for most students).

- **The ERO programs did not increase students' vocabulary scores, nor did the programs affect students' reading behaviors or their school behaviors.** The programs did not have a statistically significant impact on students' vocabulary scores at the end of ninth grade. Nor did the programs have a statistically significant effect on how often students read school-related or non-school-related texts, or on how often students use the reading strategies taught by the two programs. Similarly, impacts on student attendance and suspensions were not statistically significant, in either the program year or the following school year.

The first two study reports also examined how well the ERO programs were implemented in the study schools, as well as the extent to which the experience of students in the ERO programs compared with the literacy support received by students not selected for the programs. A key finding from these reports is that, by the end of the second year of program operation, implementation of the reading interventions, as rated through classroom observation, was well aligned with the respective program models. In addition, schools were able to offer the programs for the entire school year. In both implementation years, students in the ERO class received a greater amount of literacy support than they would have received had they not been assigned to the program.

Overview of the ERO Study

The ERO study is both a demonstration and a rigorous evaluation of two established supplemental literacy interventions that are targeted to ninth-grade students whose reading skills are two or more years below grade level as they enter high school. The purpose of the study is to evaluate these interventions' impact on students' reading comprehension skills and their academic performance as they move through high school. See Box ES.1 for a brief overview of the components of the ERO study.

Box ES.1

Overview of the Study

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

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

Research design: Within each district, high schools were randomly assigned to use either the RAAL program or the Xtreme Reading program during two school years (2005-2006 and 2006-2007). Within each high school, students were randomly assigned to enroll in the ERO class (ERO group) or to remain in a regularly scheduled elective class (non-ERO group). Because students were randomly assigned to the ERO program, the impact of the programs can be estimated by comparing the outcomes of students in the ERO and the non-ERO group. Impact estimates are regression-adjusted for the blocking of random assignment as well as random baseline differences between the ERO and non-ERO group.

Data collection: Classroom observations in the first and second semester of the school year were used to measure implementation fidelity. A reading comprehension test — the Group Reading Assessment and Diagnostic Examination (GRADE) — and a survey were administered to students in the spring of eighth grade or at the start of ninth grade prior to random assignment, and again at the end of ninth grade. School records data were collected for students' ninth-grade and tenth-grade year (and for Cohort 1's eleventh-grade year); these data include course transcripts, state test scores, attendance, and disciplinary outcomes.

Outcomes: Reading comprehension test scores, vocabulary test scores, and self-reported reading behaviors at the end of ninth grade; grade point average, credit accumulation, state test scores, school attendance and suspensions during ninth grade and in the following school year.

A Demonstration of Supplemental Literacy Interventions

The two reading programs selected for the ERO study, RAAL and Xtreme Reading, were selected for the study from a pool of 17 applicants by a national panel of experts on adolescent literacy.[11] The applicants had responded to a public request for supplemental reading

[11]For an overview of research related to RAAL, see Schoenbach, Greenleaf, Cziko, and Hurwitz (1999). For an overview of research related to Xtreme Reading and the Strategic Instruction Model, see Schumaker and Deshler (2003, 2004).

programs that would substantially improve students' reading skills. The programs share common core goals and instructional strategies, and can therefore be considered part of the same broad class of literacy intervention for struggling adolescent readers. The short-term goal of the programs is to help ninth-grade students adopt the strategies and routines used by proficient readers, to improve their comprehension skills, and to motivate them to read more and to enjoy reading. To do so, each program supports 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.

By addressing the reading needs of students, the programs also aim to improve students' performance in their high school courses and on standardized assessments, thereby helping them meet the milestones required for grade promotion and graduation. The programs seek to improve these longer-term outcomes by targeting not only students' reading skills but also their content literacy — by identifying, modeling, and explaining context-specific strategies that are most applicable in English language arts, science, and social studies texts (for example, differences in text structures). Finally, both interventions have components that promote positive behavioral norms, which in turn may improve students' attendance rates and reduce their disciplinary infractions.

Experienced, full-time English language arts or social studies teachers volunteered to teach the ERO class and were approved by ED, the districts, and the schools to teach the programs for a period of two years. During each year of the project, the programs' developers provided three types of training and technical assistance to the ERO teacher from each school: (1) a summer training institute before the start of the school year (five days in the first year of the study and three days in the second year); (2) booster training sessions; and (3) coaching visits during the school year (a minimum of two one-day visits during the first year and three two-day visits during the second year).

Each ERO teacher (one per school) was responsible for teaching four sections of the ERO class. Each section accommodated between 10 and 15 students. Classes were designed to meet for a minimum of 225 minutes per week and were scheduled as a 45-minute class every day or as a 75- to 90-minute class that met every other day during the school year. As noted earlier, the classes are supplemental, in that they replace 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 average annual cost of the programs, as implemented, was $1,931 per student. Salary expenditures represent the largest portion of this cost (72 percent). An additional 13.4 percent of the per-student cost was spent on training resources, while 5.9 percent paid for

travel to and from training activities. The remaining 8.5 percent covered equipment/supplies, other direct costs, and indirect costs.[12]

A Rigorous Impact Evaluation

The supplemental reading programs were implemented in 34 high schools from 10 school districts across the country. The districts were selected through a special grant competition organized by the U.S. Department of Education's Office of Vocational and Adult Education (OVAE). As an extension of the Smaller Learning Communities (SLCs) grant program, this competition sought to provide funding for the implementation of two supplemental ninth-grade literacy programs in selected high schools and to sustain and enhance existing SLCs in these high schools.

The ERO evaluation uses a two-level random assignment research design. First, within each district, high schools were randomly assigned to use one of the two supplemental literacy programs: 17 high schools were assigned to use RAAL, and 17 schools were selected to use Xtreme Reading. Each school implemented the same program in both the 2005-2006 and 2006-2007 school years. In the second stage of the study design, eligible students within each of the participating high schools were randomly assigned either to enroll in the ERO class (ERO group) or to take one of their school's regularly offered elective classes (non-ERO group).

Across both years of implementation, the participating high schools identified 5,595 ninth-grade students reading two to five years below grade level (an average of 82 students per school per cohort). Fifty-seven percent of these students were randomly assigned to the ERO group and 43 percent randomly assigned to the non-ERO group. Random assignment resulted in two groups of students that were similar with respect to their background characteristics and prior achievement at the start of the study.

The ERO evaluation uses a variety of data sources to measure impacts on student outcomes as well as the nature and quality of program implementation. To learn about the fidelity with which the programs were implemented, the study conducted observations of the supple-

[12]Program costs were calculated by identifying the key program inputs and resources (that is, personnel salaries and fringe benefits, training and travel costs, equipment and supplies, other direct costs, and indirect costs), and then calculating program costs based on the unit cost and quantities of these resources. Costs per student were then obtained by dividing the total cost by the number of students enrolled in the ERO programs in the study schools. Cost data were obtained from documents prepared by districts as part of their application for the Smaller Learning Community (SLC) grant, developer budget documents, teacher salary step schedules obtained from district Web sites, and the National Center for Education Statistics (NCES) Common Core of Data (CCD) Fiscal Survey (F-33 data).

mental literacy classes during the first and second semester of the school year. To measure the amount of literacy instruction received by students — as well as the contrast between the nature and quantity of the literacy services received by students in the ERO and non-ERO groups — the study team used a combination of attendance records from ERO teachers and a student survey. To measure program effectiveness, data were collected on four types of student outcomes: reading achievement (reading comprehension and vocabulary scores from the Group Reading Assessment and Diagnostic Examination, also known as the GRADE),[13] reading behaviors (frequency of reading and use of reading comprehension strategies from a student survey), academic performance in core subject areas (GPA, credits earned, and state assessment scores from school records), and school behaviors (attendance and disciplinary data from school records).

Implementation Findings

The ERO study examined three aspects of program implementation that may influence whether and by how much the reading interventions can improve student outcomes: (1) the extent to which the study schools were implementing the ERO programs as specified by the program developers (implementation fidelity), (2) the amount of ERO instruction received by students (dosage), and (3) whether the literacy services received by students in the ERO group differed in amount and type from the services received by students in the non-ERO group (service contrast). These implementation findings are discussed in the first two study reports and summarized below.

- By the end of the second year, most participating high schools' implementation of the ERO programs was well aligned with the program models.

During each year of the project, observers from the study team used a structured classroom observation protocol to examine whether the study schools were implementing the ERO programs as specified by the program developers.[14] The overall implementation of the ERO program in a given school was classified as "well aligned," "moderately aligned," or "poorly

[13]The GRADE is a norm-referenced, research-based reading assessment that is used widely to measure performance and track the growth of an individual student and groups of students. The average score on the reading comprehension or vocabulary subtests is 100 for a nationally representative group of students at the end of their ninth-grade year. The national standard deviation of scores for both tests is 15. For more information, see American Guidance Service (2001a, 2001b).

[14]The analysis of implementation fidelity is based on three field research visits to each of the 34 high schools — one during the second semester of the first implementation year, and one in each of the first and second semesters of the second implementation year.

aligned," based on observers' ratings of how reflective the ERO classroom's learning environment (classroom climate) and comprehension instruction (the teacher's use of ERO instructional strategies) were of the behaviors and activities specified by the developers. At the spring site visit in the second year of the study, the ERO courses at 26 of the 34 schools (76 percent of schools) were categorized as being well aligned with the program models.

- In the second year of the study, schools operated the ERO programs for the entire school year, and student participation was close to what was intended.

To measure the amount of ERO instruction received by students (dosage), the study team collected data on the duration of the ERO classes as well as the frequency with which students attended the ERO classes, from ERO teachers' attendance records. On average, students in the ERO group attended 79 percent of scheduled ERO classes, and they received 98 hours of ERO instruction during the school year, or 11 hours per month, which is the same monthly average as in the first year of implementation. The programs were designed to meet for a minimum of 225 minutes per week (15 hours per month), and the schools were able to offer the ERO classes an average of 14 hours per month. Thus, the 11 hours per month of ERO instruction received by students represents 79 percent of the amount offered and 73 percent of the amount intended.

- In both implementation years, students in the ERO group reported a higher frequency of participation in supplemental literacy services than students in the non-ERO group.

For a program to have an impact on outcomes, it needs to provide services that differ (for example, in quality, nature, frequency) from the services students would have otherwise received; this difference is known as the "service contrast." The ERO study team collected data to better understand what types of literacy services were received by the non-ERO students and the extent to which such supports may have reduced the service contrast between the ERO and non-ERO group. A student survey (administered at the end of ninth grade) was used to measure ERO and non-ERO students' participation in different types of supplementary literacy support activities (including the ERO class).[15] These data confirm that, as expected, students in the ERO group attended a greater number of school-based literacy classes annually than students in the non-ERO group. On average, the ERO students reported attending 52 more literacy class sessions than the non-ERO students in the first year of implementation and 58 more sessions in the second year.

[15]Although students in the ERO group were not specifically instructed to include ERO classes in their survey response, it is highly likely that their responses include this class.

Impact Findings

The previous two study reports — which focused on the effect of the programs on students' reading outcomes in ninth grade — showed that the ERO programs improved students' reading comprehension scores. Students' reading achievement at the end of ninth grade was measured using the reading comprehension and vocabulary subtests in the GRADE.

- The ERO programs improved students' reading comprehension scores over the course of ninth grade.

Figure ES.1 shows the impact of the ERO programs on students' GRADE reading comprehension test scores in standard score units at the end of ninth grade.[16] The impact on students' reading comprehension scores was 0.9 standard score point (effect size = 0.09).[17] As seen in the figure, students in the ERO group started their ninth-grade year with an average standard score of 85.3, which corresponds, approximately, to a grade equivalent of 5.0 (the start of fifth grade) and a reading level at the fifteenth percentile for ninth-grade students nationally. In the spring of ninth grade after enrolling in the program, the ERO group's average score was 90.1 points (twenty-fifth percentile nationally), which means that these students' reading scores increased by 4.9 points over the course of ninth grade. In contrast, had these students not been assigned to the ERO class, their reading scores would have increased by 4.0 standard score points during the year (the estimated growth of the non-ERO group), to 89.2 points or the twenty-third percentile nationally. The impact of the ERO programs is the difference between the growth in test scores of the two groups (0.9 standard score point) and represents an 23 percent improvement over and above the growth that the ERO group would have experienced if they had not had the opportunity to attend the ERO classes (4.0 points).

Despite this impact, ERO students' reading skills were still below grade level at the end of the program. As shown by the solid line at the top of Figure ES.1, students with a score of 100 points on the GRADE reading comprehension test at the end of ninth grade are considered to be reading at grade level. In contrast, students in the ERO group had an average standard score of 90.1 at the end of ninth grade, which means that ERO students were nearly 10 points below the national average, or almost four years below grade level. More concretely, 77 percent of students in the ERO group would still be eligible for the ERO programs were these programs

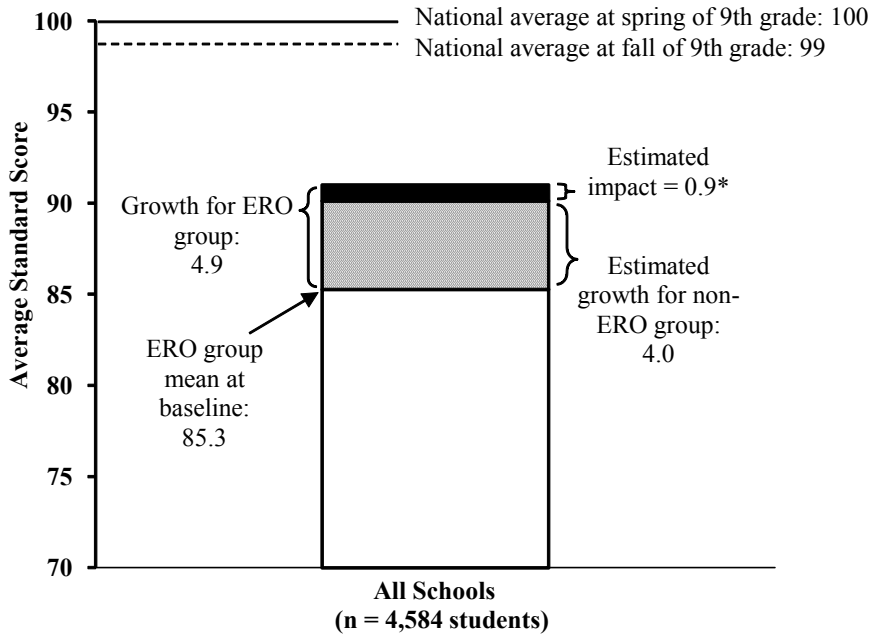
[16]Impact estimates are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort and for random differences between the ERO and non-ERO groups in their baseline characteristics and prior achievement.

[17]The impact on students' reading comprehension scores does not differ by a statistically significant amount between the two implementation years (effect size = 0.09 for students in the first cohort and 0.08 for students in the second cohort).

The Enhanced Reading Opportunities Study

Figure ES.1

Impacts on Reading Achievement, GRADE Respondent Sample



SOURCE: MDRC calculations from the Enhanced Reading Opportunities GRADE assessment, administered at the end of ninth grade (spring 2006 for Cohort 1 and spring 2007 for Cohort 2).

NOTES: The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline and whether a student was overage for grade at random assignment. 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 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.

Rounding may cause slight discrepancies in calculating sums and differences.

[a]The national average for standard scores 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.

again made available to them (because they scored two or more years below grade level at the end of their ninth-grade year).

This report presents new findings related to the ERO programs' impact on students' academic performance and their behavioral outcomes, in ninth grade (the "program year") and in the subsequent school year (tenth grade for most students). School records provided by the study districts included information on students' GPA in core subject areas (English language arts, social studies, and science), the number of credits earned by students in these subject areas, and students' scores on the tests mandated by their state.

- The ERO programs had a positive impact on students' GPA and credits earned while students were in the program (ninth grade), but these impacts did not persist into the following school year.

As shown in the top panel of Figure ES.2, students in the ERO group had a GPA of 1.60 points during their ninth-grade year (out of a maximum of 4 points), while students in the non-ERO group had a GPA of 1.53 points, which means that both groups of students had a D average during the program year. However, the GPA of students in the ERO group was statistically higher than that of students in the non-ERO group (a difference of 0.06 point; effect size = 0.07). This improvement is such that ERO students were 13 percent closer to achieving C average (2.0 points), which is an important milestone associated with a higher likelihood of graduating from high school.[18,19]

As shown in the top panel of Figure ES.3, the ERO programs also helped students accumulate a greater number of credits in core subject areas. By the end of ninth grade, students in the ERO group had earned 2.99 credits (or 21.4 percent of the core credits that they need to graduate), while students in the non-ERO group had earned 2.91 credits (20.9 percent of the core credits needed to graduate). To put these findings into perspective, consider that at the end of ninth grade, students should have earned 25 percent of the core credits required for graduation in order to be "on track" to graduate. Therefore, by the end of the program year, neither the ERO group nor the non-ERO group was "on track" to graduate in four years. However, students in the ERO group had earned a greater percentage of required core credits than students in the non-ERO group (a difference of 0.6 percentage point; effect size = 0.06). This improvement is

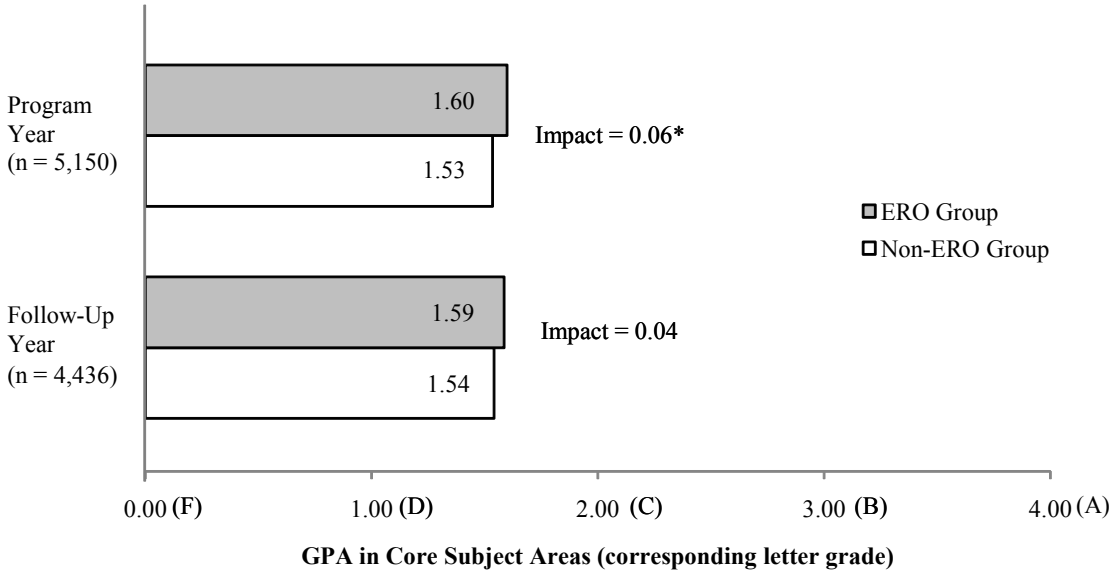
[18]Based on Allensworth and Easton (2007), graduation rates in the Chicago Public Schools are 36 percentage points higher for students with a C average (2.0) than for students with a D average (1.0). A grade of C is also used by a majority of the study schools to determine "average" or "satisfactory" achievement in a given course.

[19]The value of 13 percent was calculated by dividing the impact (0.06 points) by the amount by which students in the ERO group are below a C average (0.47 points).

The Enhanced Reading Opportunities Study

Figure ES.2

Impacts on Grade Point Average (GPA), School Records Sample



SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group values are the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment.

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.

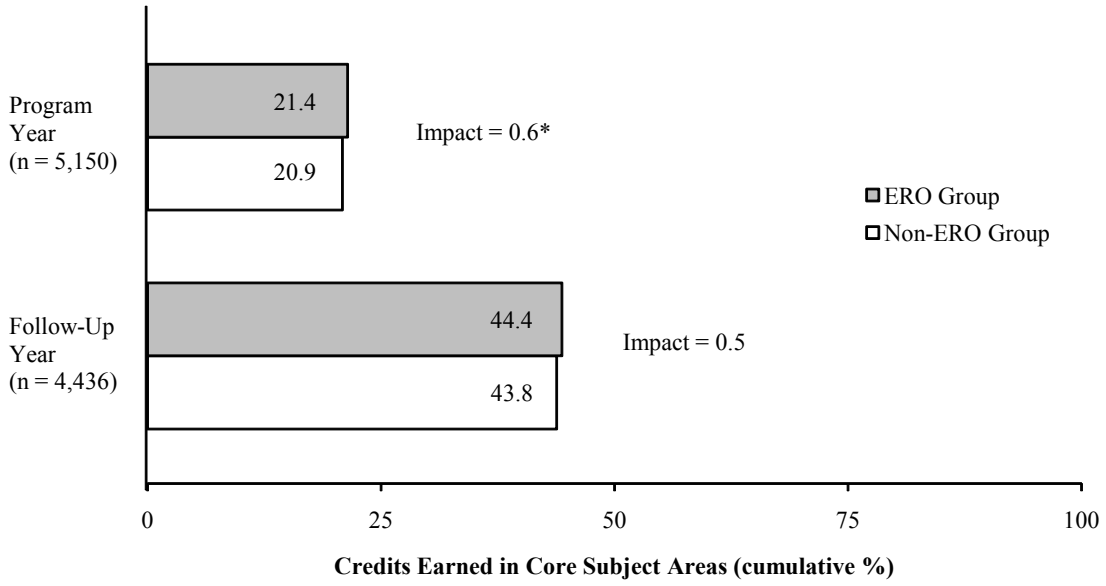
Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The Enhanced Reading Opportunities Study

Figure ES.3

Impacts on Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample



SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment.

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 cumulative number of credits earned is scaled as a percentage of the total number of core credits required for graduation in a student's district.

statistically significant, and its magnitude is such that students in the ERO group were 15 percent closer to being "on-track" to graduate as a result of the ERO program.[20,21]

The ERO programs did not affect students' GPA or credits earned in core classes in the school year following their participation in the ERO programs. The lower set of bars in Figure ES.2 shows the average GPA in core courses for ERO students and non-ERO students during the follow-up year (1.59 and 1.54 points, respectively).[22] The estimated impact of 0.04 point is not statistically significant. Nor did the ERO programs have a statistically significant impact on credits earned by the end of the following school year (lower set of bars in Figure ES.3). At the end of the follow-up year, on average, non-ERO students had accumulated 43.8 percent of the credits they need to graduate, while ERO students had accumulated 44.4 percent on average.[23] The estimated impact of 0.5 percentage point is not statistically significant. Therefore, it cannot be concluded that the programs had improved students' GPA or credit accumulation by the end of the follow-up year.

- The ERO programs had a positive impact on students' performance on state tests in English language arts and mathematics in ninth grade. Impacts on state test scores in the following school year are not statistically significant.

Table ES.1 shows findings for the estimated impact of the ERO programs on students' scores on the tests mandated by their state. In the subset of schools where state tests are administered in the ninth grade, the programs had a positive impact on students' performance on state tests in English language arts (effect size = 0.11) and mathematics (effect size = 0.07). However, impacts on state test scores in the follow-up year were not statistically significant.

- The ERO programs did not have a statistically significant impact on students' reading vocabulary scores or on their reading behaviors, nor did it affect their school behaviors.

---

[20]The value of 15 percent was calculated by dividing the impact (0.6 percentage point) by the percentage of credits needed by the ERO group to attain 25 percent of their core course credits (4.1 percentage points).

[21]Viewed otherwise, the ERO programs had an estimated impact of 0.08 on the number of core credits earned by students (the ERO group earned 2.99 credits on average in the program year, while the non-ERO group earned 2.91 credits). One credit represents a full-year course, so students in the ERO group earned an additional 8 percent of a full-year course credit relative to the non-ERO group.

[22]Note that the GPA measure is not cumulative and includes only students' grades in core courses during the given school year.

[23]The measure of credit accumulation is defined cumulatively, in order to capture a student's progress toward graduation. Thus, credit accumulation at the end of the follow-up year includes credits earned during the program year *and* the follow-up year.

The Enhanced Reading Opportunities Study

Table ES.1

Impacts on State Test Scores (Standardized), School Records Sample

Outcome	Number of Students	ERO Group	Non-ERO Group	Estimated Impact Effect Size[a]	P-Value for Estimated Impact
All schools					
Program year					
English language arts (ELA)	2,244	0.11	0.01	0.11*	0.003
Social studies	952	0.07	0.01	0.06	0.343
Science	2,348	0.07	0.01	0.06	0.166
Math	2,668	0.08	0.01	0.07*	0.045
Follow-up year					
English language arts (ELA)	2,408	0.00	-0.01	0.01	0.705
Social studies	2,237	-0.04	0.01	-0.04	0.262
Science	2,661	0.00	0.02	-0.02	0.649
Math	2,537	-0.03	0.02	-0.04	0.265

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by their school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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 numbers of students reported in this table are for students in the school records sample who have state test scores for a given subject area in the relevant year. A student may have taken more than one test in a given subject area in more than one year. If a student wrote a specific test more than once, only his or her first score is used.

Rounding may cause slight discrepancies in calculating sums and differences.

[a]State test scores are standardized by district, follow-up year, and cohort, using the means and standard deviation of the non-ERO group.

The previous two study reports showed that the ERO programs did not have a statistically significant impact on students' scores on the GRADE vocabulary subtest at the end of ninth grade. Nor did the programs have a statistically significant effect on the frequency with which students read inside or outside of school, or on the extent to which they use the different kinds of reading strategies taught by the ERO programs, based on a student survey administered at the end of ninth grade. The new findings in this report show that, similarly, the programs did not have a statistically significant effect on students' attendance rate, nor did they affect whether students were suspended, in either ninth grade or the following school year. Information on these student behavioral outcomes was available from records provided by the study districts.

- There is no conclusive evidence that the programs were more effective for one subgroup of students than another.

In this report and in prior reports, the impact of the reading programs on reading and high school outcomes was estimated for subgroups of students defined by their baseline reading achievement level, whether or not they were overage for grade at the start of ninth grade, and whether or not a language other than English is spoken in their home. Based on these analyses, it cannot be concluded that the ERO programs' impacts on students in a given subgroup were different from their impacts on students in the other subgroup(s) in that category.

- Each of the two ERO programs had positive impacts for students during the program year, but the statistical certainty (significance) of the impacts varied by outcome.

Because the primary goal of the ERO study is to estimate the impact of full-year supplemental reading programs, the findings in this report focus on the combined impact of the two reading programs together (RAAL, Xtreme Reading). However, in order to contextualize the overall impact findings, program-specific impacts are also examined in this report. Impact findings for each of the two reading programs separately tell a similar story to the pooled findings for both programs together. As shown in Table ES.2, both programs had a positive effect in ninth grade on GPA (statistically significant) and credit accumulation (not statistically significant). For both programs, these two impacts were similar in magnitude to the overall impacts for both programs pooled together. Neither of the two programs had a positive impact on academic performance in the follow-up year. For the GRADE reading comprehension test scores in ninth grade pooled across cohorts, the impact of RAAL was 1.2 standard score points (effect size = 0.12) and statistically significant, and the impact of Xtreme Reading was 0.6 standard score point (effect size = 0.05) and not statistically significant. The difference in the

The Enhanced Reading Opportunities Study

Table ES.2

Impacts on Grade Point Average (GPA) and Credit Accumulation in Core Subject Areas (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools					
Program year					
GPA in core subject areas	1.60	1.53	0.06 *	0.07 *	0.002
Credits earned in core subject areas (%)	21.4	20.9	0.6 *	0.06 *	0.017
Sample size	2,937	2,213			
Follow-up year					
GPA in core subject areas	1.59	1.54	0.04	0.05	0.061
Cumulative credits earned in core subject areas (%)	44.4	43.8	0.5	0.03	0.212
Sample size	2,542	1,894			
Reading Apprenticeship schools					
Program year					
GPA in core subject areas	1.54	1.47	0.07 *	0.07 *	0.019
Credits earned in core subject areas (%)	20.5	19.9	0.6	0.06	0.072
Sample size	1,468	1,095			
Follow-up year					
GPA in core subject areas	1.55	1.52	0.03	0.03	0.392
Cumulative credits earned in core subject areas (%)	42.6	42.1	0.4	0.03	0.486
Sample size	1,285	927			
Xtreme Reading schools					
Program year					
GPA in core subject areas	1.66	1.60	0.06 *	0.06 *	0.038
Credits earned in core subject areas (%)	22.3	21.8	0.5	0.05	0.138
Sample size	1,469	1,118			
Follow-up year					
GPA in core subject areas	1.62	1.56	0.06	0.06	0.068
Cumulative credits earned in core subject areas (%)	46.2	45.5	0.7	0.04	0.266
Sample size	1,257	967			

(continued)

Table ES.2 (continued)

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by their school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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 sample sizes reported in this table are for students with school records data in the relevant year. Because students may not have earned grades in all core subject areas in a given year, sample sizes differ for impacts in the specific core subject areas.

Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The cumulative number of credits earned is scaled as a percentage of the total number of credits (core or subject-specific) required for graduation in a student's district.

reading comprehension impacts of the two programs is not significant, and thus it cannot be concluded that one program was more effective than the other. Across all other student outcomes measured in this report (16 in total),[24] individual program impacts were statistically different from one another for only two outcomes (state test scores in science in the program year and social studies tests in the follow-up year).

Poststudy Implementation of the ERO Programs

After the two years of implementation required by the study, ERO schools and districts continued to receive SLC grant funding, but they were free to decide whether to continue the

[24]This includes eight types of outcome (GPA, credit accumulation, attendance, suspensions, and state test scores in each of the four core subject areas) measured in two follow-up years.

ERO programs or to use the funds to improve other aspects of their SLCs. Interviews were conducted with school-level staff to find out whether the study schools continued to use the ERO programs after the study-required implementation period, and if so, in what ways (if any) the programs have been modified to fit local circumstances. The study team was able to interview school-level staff from 30 of the 34 study schools.

- Fourteen high schools (47 percent of the interviewed schools) continued to offer the ERO program after the end of the study-required implementation period. Schools that continued to use the ERO programs modified them.

Deviations from the implementation conditions required by the study aimed at increasing the number of students served by the programs (for example, increasing class size [10 schools], or serving students other than ninth-graders [seven schools]) and/or lengthening their duration [seven schools]). Also, eight schools modified the content of the programs and nine schools reduced the levels of professional development and technical assistance provided to teachers. The modifications that these schools made to the programs may alter the programs' effectiveness relative to the impact findings.

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.[1] Of these students, nearly half do not exhibit even partial mastery of the knowledge and skills that are fundamental to proficient work at grade level.[2] These limitations in literacy skills are a major source of course failure, high school dropout, and poor performance in postsecondary education.[3] 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.[4]

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.[5] The ERO demonstration involves 34 high schools from 10 school districts across the country that implemented one of two supplemental literacy programs: Reading Apprenticeship Academic Literacy (RAAL), designed by WestEd, or Xtreme Reading, designed by the University of Kansas Center for Research on Learning. 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 districts that participated in the study were selected through a special grant competition organized by the U.S. Department of Education's Office of Vocational and Adult Education (OVAE).[6] Anticipating that there would be

[1]The NAEP is a nationally representative assessment of student achievement overseen by the National Center for Education Statistics (http://nces.ed.gov/nationsreportcard/).

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

[3]Carnevale (2001); Kamil (2003); Snow and Biancarosa (2003).

[4]Biancarosa and Snow (2004).

[5]The ERO study is known more formally as "An Evaluation of the Impact of Supplemental Literacy Interventions in Freshman Academies."

[6]For 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 (SLC) 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 SLC initiatives under way in the districts. In 2006, responsibility for the SLC initiative and for the special ERO grants was moved from OVAE to Office of Elementary and Secondary Education (OESE).

first-year start-up challenges, the U.S. Department of Education supported a demonstration and evaluation of these programs that included two years of program operations in the study schools (2005-2006 and 2006-2007 school years), resulting in two cohorts of participating ninth-grade students. MDRC — a nonprofit, nonpartisan social policy research organization — conducted the evaluation in partnership with the American Institutes for Research (AIR) and Survey Research Management (SRM).

The ERO study assesses the impact of the two supplemental literacy programs on students' reading comprehension skills and on their general performance in high school. The first two reports for the study focused on measuring how well the programs were implemented and evaluated the programs' impact on students' reading achievement and behaviors.[7] This report — the final of three reports from this evaluation — examines the impact of the programs on students' high school outcomes during the year in which they were enrolled in the ERO programs (ninth grade), as well as the following school year (tenth grade for most students). The academic outcomes that are the focus of this report include students' course grades, their accumulation of the credits they need to graduate, as well as their scores on state standardized assessments. The report also examines the programs' effect on students' school behaviors, which include attendance and disciplinary infractions. These school-based outcomes — course grades, credit earning, state assessment scores, attendance, and disciplinary infractions — represent measures typically used by schools and districts to assess how their students are doing.[8] In addition, as discussed more fully later in this chapter, analyzing the impacts on these outcomes tests the subsequent step in a theory of change. This theory posits that addressing the reading needs of students will help them do better not only on reading assessments, but also in their schoolwork more broadly. To set a foundation for the presentation and discussion of the analyses of the school-based outcomes, this report reviews the findings from the first two reports. Thus, this report provides comprehensive coverage of what has been learned from this evaluation.

The remainder of this chapter provides background on the ERO demonstration, including descriptions of the two interventions selected for the project, their implementation in study schools in 2005-2006 and 2006-2007, and an overview of the research design used to assess impacts on key student outcomes.

---

[7]See Kemple et al. (2008) and Corrin et al. (2008). Chapter 3 is a review of the findings from these first two reports — the impact findings on reading test scores and reading behaviors as well as the implementation findings from research on the operation of the two programs.

[8]Throughout the report there will be references to "high school outcomes" or "school-based outcomes," which refer to outcomes from school records data — that is, outcomes from data regularly collected by the schools participating in the study. The reading achievement and reading behaviors outcomes that were the focus of the first two reports are considered "study-based outcomes," as they were measured using instruments administered specifically for this evaluation.

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 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 the Institute of Education Sciences (IES) is overseeing the evaluation effort. Following is a brief overview of the demonstration and evaluation components of the ERO study.

Overview of the 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 — RAAL, designed by WestEd, and Xtreme Reading, designed by the University of Kansas Center for Research on Learning — were selected for the study from a pool of 17 applicants by a national panel of experts on adolescent literacy.[9,10]

The two programs are similar in terms of their instructional approach and overarching goals and can therefore be considered part of the same broad class of supplemental literacy intervention geared at struggling adolescent readers. The core objectives of both programs are to help ninth-grade students adopt the strategies and routines used by proficient readers, improve their comprehension skills, and motivate them to read more and to enjoy reading. Each program supports 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. While both programs seek to improve students' reading skills in general, they address content literacy as well — identifying, modeling, and explaining context-specific strategies that are most applicable in English, science, and social studies texts (for example, differences in text structures).[11] The RAAL program has units that specifically target content areas (for example, "Reading History" and "Reading Science and Technology"); the Xtreme

[9]For an overview of research related to RAAL, see Schoenbach, Greenleaf, Cziko, and Hurwitz (1999). For an overview of research related to Xtreme Reading and the Strategic Instruction Model, see Schumaker and Deshler (2003, 2004).

[10]For further detail about the selection process, see Appendix A.

[11]While both programs also address reading for math courses, this content area receives less attention than the other three core content areas.

Reading program draws on texts from different disciplines, and teachers encourage students to bring in texts and work from other courses to use in the literacy class.

The ERO programs' theory of change is depicted in Figure 1.1. As seen here, the short-term goal of the ERO programs is to improve students' reading achievement. In turn, this improvement in reading skills is hypothesized to help students better understand the material taught in their courses, thereby improving their performance in their courses and on high-stakes tests. This improvement in their academic performance should increase students' accumulation of course credits, thereby helping them meet the milestones required for grade promotion and, ultimately, for obtaining their high school diploma. As shown in Figure 1.1, the ERO programs may also affect students' academic outcomes through a secondary pathway. While the primary focus of the reading programs is on the teaching of reading strategies, both interventions also have components that promote positive school-related behavioral norms. Thus, the ERO programs may also improve students' attendance rate and reduce their behavioral problems. This in turn may facilitate students' ability to do better in their classes and progress toward graduation. (For example, regular attendance is sometimes a requirement for passing a course.)[12]

Implementation of the Programs and Teacher Training

Both RAAL and Xtreme Reading 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.[13] Each program is a full-year course 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.[14]At each of the 34 high

[12]Appendix A contains detailed descriptions of RAAL and Xtreme Reading, including a table that lists the key components of each program. The one area where the two programs differ is with respect to the flexibility allowed to teachers. Implementation of RAAL is guided by the concept of "flexible fidelity" — that is, while the program includes a detailed curriculum, teachers have flexibility in how they include various aspects of the RAAL curriculum in their day-to-day teaching activities, but have been trained to do so such in a way that maintains the overarching 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.

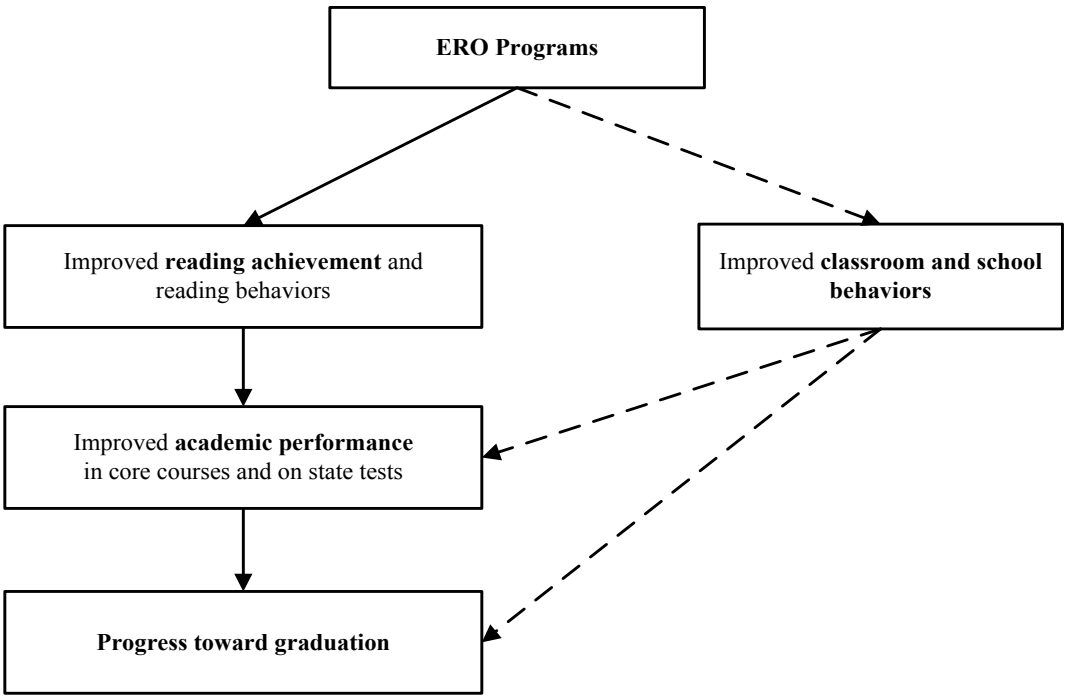
[13]For a discussion and examples of how ERO classes fit into schools' existing schedule models and how ERO students' class schedules compared with non-ERO students' class schedules, please see the section "The ERO Classes and Student Schedules" in Appendix A and Table A.2.

[14]Both reading programs were adapted by the developers in two ways for this study from already existing comprehensive high school literacy programs. First, the developers modified their comprehensive programs for implementation as a supplemental elective course. In addition, the developers tailored their professional development and coaching strategies to meet the needs of high school teachers selected by participating schools who, by design, did not have reading instruction credentials.

The Enhanced Reading Opportunities Study

Figure 1.1

The ERO Programs' Theory of Change



schools participating in the study, an experienced, full-time English language arts or social studies teacher volunteered to teach the programs for a period of two years.[15] Training and technical assistance were delivered to the ERO teachers in several ways: ERO teachers attended summer training institutes held prior to each year of implementation. In addition, RAAL teachers attended two two-day booster training sessions, and Xtreme Reading teachers attended one two-day booster training each year. The program developers also provided onsite support to the ERO teachers through instructional coaching visits to the schools both years. RAAL teachers also received ongoing support through access to a special online listserv that the developer set up for the project. District program coordinators were invited to observe the trainings to familiarize them with the programs in case they had to provide technical assistance or other support to ERO teachers. Of the 34 ERO teachers who began teaching the ERO courses in the first year of implementation, 25 returned and taught all of the second year of the program. Of the nine original teachers who did not teach two full years of the program, two were replaced

[15]For descriptive information about the teachers who taught the ERO courses, please see Appendix A.

during the first year and seven were replaced after the conclusion of the first year. The nine replacement teachers taught all of the second year of the program.[16,17] Figure 1.2 provides a timeline for both years of program implementation, including teacher selection, teacher training, and teacher replacement.[18]

A Rigorous Impact Evaluation

The ERO study addresses questions about the impacts of the ERO programs on student outcomes, as well as questions related to the implementation of the programs. The first two reports from the study addressed the following questions:

- What were the impacts of the two supplemental literacy interventions, together and separately, on ninth-grade students' reading skills and reading behaviors?
- How and how well were the two interventions implemented? How much did implementation vary across participating high schools?

The findings presented in those reports will be reviewed in Chapter 3 of this report. This report also presents new findings in response to the following research question:

- What are the impacts of the two supplemental literacy interventions, together and separately, on other academic outcomes, such as performance in academic courses, progress toward graduation, and achievement on high-stakes standardized tests, during both the year in which the students were participating in the ERO programs (the program year) and the following year?

Additionally, the reports discuss how the impacts of the programs vary (or not) for different subgroups of students and examine associations between implementation fidelity and quality and program impacts (or lack of impacts) on student outcomes.

To answer these questions, the ERO evaluation uses 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 process ensured that each program developer was assigned a fair draw of schools in which to implement their programs. In the second

[16]Teachers had to be replaced because of mobility and turnover (that is, teachers moving, transferring, or retiring) or poor performance based on school or district teacher evaluation processes exogenous to the evaluation.

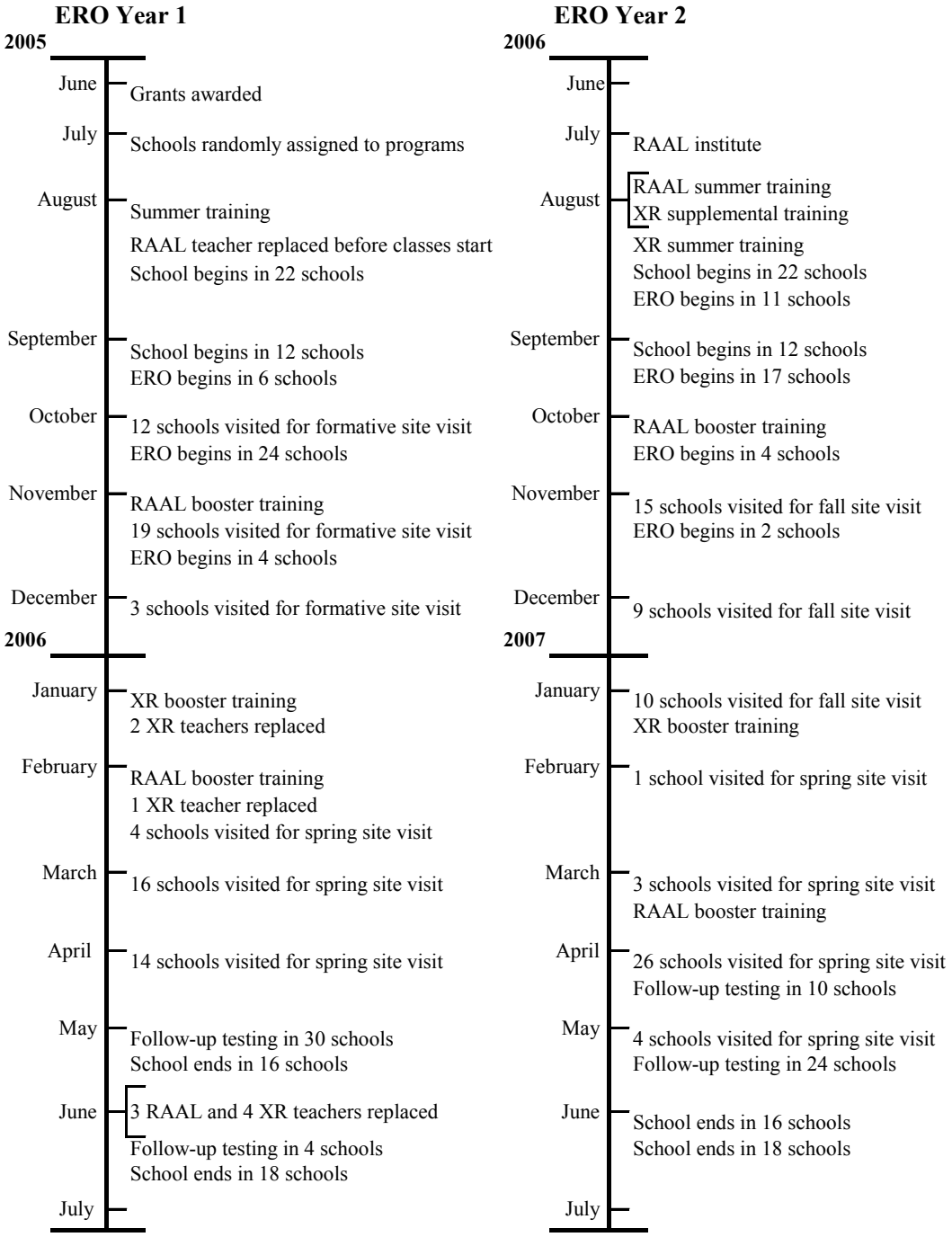
[17]Replacement teachers received additional professional development support during the second year and the preceding summer, compared with the returning teachers. This additional support was designed to give them a level of support similar to that ERO teachers received in the first year of implementation.

[18]More detailed information about teacher training is available in Appendix A.

The Enhanced Reading Opportunities Study

Figure 1.2

Study Timeline: Implementation Years



SOURCES: MDRC and AIR project records and documentation.

stage of the design, eligible and appropriate students within each of the participating high schools were randomly assigned to either enroll in a supplemental literacy class (ERO group) or to remain in one of the regular elective classes available to ninth-grade students (non-ERO group). This feature of the design is possible because there were more eligible students in each high school than 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 for the 2005-2006 school year and one cohort that was enrolled in the study for the 2006-2007 school year.

As noted earlier, the first two reports for this study examined impacts on participating students' reading achievement and reading behaviors for each of the two years of program implementation, while this report focuses on impacts on school performance in ninth and tenth grades. The ERO evaluation taps a variety of data sources to measure these outcomes and to assess the fidelity of program implementation. These data sources include a standardized reading assessment, a student questionnaire, student records from participating districts, classroom observations, and class schedule information.

There are two important points to note when interpreting the findings produced by this study. First, the ERO study was designed primarily to test the combined impact of Xtreme Reading and RAAL as a broad class of reading intervention. As explained earlier in this chapter, Xtreme Reading and RAAL share many common instructional principles and strategies, so they can be evaluated together as full-year supplemental literacy courses targeted at struggling adolescent readers. The ERO study's design also makes it possible to evaluate the effectiveness of each of these two programs separately in order to provide context for interpreting the combined impact of the two programs. However, given the similarities between the two programs, the purpose of the ERO study is not to evaluate the differential impact of the two interventions. As described earlier, the two programs were randomized to schools, but the purpose of this randomization was primarily to ensure that each program developer was assigned a fair draw of schools in which to implement their programs, rather than to test for a differential impact between the two interventions. For this reason, the statistical model chosen for the impact analysis does not use the school-level randomization feature of the research design; moreover, the study's design allows only for the detection of relatively large differences in impacts between the two programs. Although tests of the difference in impacts between the two programs are conducted in this report, these tests are used only as a means of verifying that the programs' impact does not differ by a statistically significant amount and that they can be

pooled together as a class of intervention. The results of these tests are not appropriate for making inferences about the true differential impact of the two programs.[19]

It is also important to note that the 34 schools in this study 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 most applicable and relevant to similar districts and high schools that are struggling to meet the needs of ninth-graders who lack the literacy skills required for academic success.

Overview of This Report

As noted earlier, this report focuses on the ERO programs' impact on students' academic performance and their school behaviors. However, before proceeding to these findings, the next chapter provides a broad overview of the ERO study's design, which includes a discussion of various analytic issues relevant to this report as well as the two prior reports. Chapter 3 reviews what was learned in the first two reports, in terms of the operation of the supplementary literacy interventions during their two years of implementation, as well as their impacts on the more proximal outcomes of students' reading test scores and reading behaviors. Chapter 4 then presents findings from new analyses of school records data, pertaining to the impact of the programs on students' academic performance and behavioral school outcomes. Chapter 5 presents findings from nonexperimental analyses whose purpose is to further understand and explore the pattern of impacts on the student outcomes in this study. Chapter 6 shares information intended to help practitioners and policymakers who are considering the implementation of supplemental literacy courses for ninth-graders in their schools or districts. In particular, the chapter presents information about program costs and whether and how literacy programming continued in the ERO high schools following the required implementation of the ERO classes in 2005-2006 and 2006-2007.

[19] As will be described in Chapter 2, the statistical model chosen for the impact analysis does not account for the school-level randomization feature of the research design, which means that it cannot be used to test for the true differential impact of the two programs.

Chapter 2

Study Design: Sample, Measures, and Analysis

This chapter provides a comprehensive overview of the research design of the Enhanced Reading Opportunities (ERO) study. The chapter begins by describing the schools and students participating in the evaluation. This is followed by a discussion of the data that were collected to measure student outcomes and to assess program implementation. The chapter then describes the samples of students and analytic methods used to assess program impacts in this report and in the two previous reports. The following key points are discussed:

- In the first stage of the study design, 34 high schools from 10 school districts were selected for the study and were randomly assigned to use one of two supplemental literacy programs — Reading Apprenticeship Academic Literacy (RAAL) or Xtreme Reading. The resulting two groups of schools were similar on a range of characteristics.

- In the second stage of the study design, eligible and consenting ninth-grade students in each school were randomly assigned either to enroll in an ERO class ("ERO group") or to remain in one of the regular elective classes available to ninth-grade students ("non-ERO group"). Two cohorts of ninth-grade students participated in the study, one for each school year in which the programs were implemented.

- Across both years of implementation, the study sample includes 5,595 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 while the remaining 43 percent were assigned to a non-ERO control group and continued in a ninth-grade elective class.

- Impacts on reading achievement — presented in the previous two study reports and reviewed in Chapter 3 of this report — are based on the sample of students who completed a reading assessment at the end of the program year (ninth grade). Across both years of implementation, a total of 4,584 study participants (representing 82 percent of the full study sample) completed the reading assessment.

- Impacts on high school outcomes — presented in Chapter 4 of this report — are based on students with school records data. A total of 5,150 students (representing 92 percent of all students in the study sample) have school records data in the program year (ninth grade). A total of 4,436 students (representing 79 percent of the study sample) also have school records data in the following year (tenth grade for most students).

- Among students with school records or reading achievement test scores, overall differences in background characteristics and prior achievement between the ERO and non-ERO groups are not statistically significant. This confirms that, for all analysis samples used in this study, random assignment resulted in two groups of students whose characteristics are statistically equivalent in expectation at baseline and that differences in outcomes between the ERO and non-ERO groups can be attributed to the impact of the supplemental reading programs.

- Statistical power calculations indicate that, across both cohorts of study participants, the study can detect impacts (in effect size) on high school outcomes as small as 0.06 in the program year and 0.07 in the follow-up year for the two supplemental reading programs pooled together (RAAL and Xtreme Reading). For each of the two reading programs separately, the study can detect impacts of 0.08 in the program year and 0.10 in the follow-up year. For reading achievement outcomes (which are the focus of prior reports), the study can detect impacts as small as 0.08 for both programs pooled together and impacts of 0.11 for each program separately.

Schools in the Study

The school districts that participated in the ERO 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).[20] As an extension of the Smaller Learning Communities (SLCs) grant program, this competition sought to provide funding for the implementation of two supplemental ninth-grade literacy programs in selected high schools and to sustain and enhance existing SLCs in these high schools.

[20]U.S. Department of Education (2005).

In June 2005, ED selected 10 grantee school districts encompassing 34 high schools from a pool of 33 applicant districts.[21] The 10 grantee districts include 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 received 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 resources.

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 use the RAAL program, and half were randomly assigned to use the Xtreme Reading program. High schools in the study implemented the same program for two school years (2005-2006 and 2006-2007).

The random assignment of schools to programs makes it possible to draw valid inferences about the differential impact of the two programs. By randomly assigning schools to one of the two supplemental literacy interventions, the study ensured that the two interventions were implemented in groups of schools that were similar on average. As a result, any difference in impacts that may emerge between the two groups of schools can be attributed to differences in the effectiveness of the two programs. Had districts or developers been allowed to choose one of the two interventions, the two literacy programs could potentially have been implemented in two very different groups of schools. This is because the decisions of districts and developers could have been based on any number of characteristics that might have made one school a better candidate than another for a successful implementation of the program. For example, developers may have favored schools with a higher level of readiness for their program or schools with fewer existing reading supports for their students. Similarly, schools may have selected the literacy intervention that they believed would be most appropriate or effective in their school. While these may be sound choices from a practitioner's or a developer's perspective, the resulting differences in school context cannot be measured and would present a threat

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

to the inference that one of the programs is more effective than the other. The randomization of schools to interventions ensures that the difference in impacts between the two groups of schools is due to a difference in their effectiveness and not to a difference in the characteristics of the schools in which the programs are operating.

Characteristics of Schools Selected for the ERO Project

Table 2.1 presents the characteristics of the 34 high schools participating in the ERO study in the year the OVAE grant was awarded (2004-2005).[22] Overall, ERO programs were implemented in schools located predominantly in large and midsize 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.[23] 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 RAAL and the schools assigned to use Xtreme Reading. RAAL schools have slightly lower promoting power and somewhat higher percentages of students eligible for Title I services, whereas the Xtreme Reading schools have slightly higher percentages of students eligible for free or reduced-priced lunch. However, none of the differences between RAAL and Xtreme Reading schools are statistically significant (p-value less than or equal to 5 percent).

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

[22]The characteristics of these schools were also examined for the school years in which the ERO programs were implemented (2005-2006 and 2006-2007). It was found that the schools had changed very little in the latter years; in addition, there was still a high degree of similarity between schools implementing RAAL and Xtreme Reading.

[23]Balfanz 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 U.S. (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 1 (%)	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, NCES 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 in the school year that the OVAE grant was awarded (2004-2005).

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.

1,000 students in grades 9 through 12, and did not select students based on past achievement or performance. This national sample of 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.

Students in the Study

The *target population* for the evaluation (as well as the reading interventions) includes students reading two to five years below grade level. However, due to difficulties recruiting a sufficient number of students in the target population, the eligibility criteria were ultimately expanded to also include students reading one to two years below grade level. Hence, the eligible population for the program is comprised of students reading one to five years below grade level.[24]

Students' eligibility for the program was determined using the Group Reading Assessment and Diagnostic Examination (GRADE). During the first study year (2005-2006), testing and recruitment took place in the fall of students' ninth-grade year. In the second study year (2006-2007), most testing and recruitment took place in the spring of students' eighth-grade year.[25] Students scoring one to five years below grade level on the GRADE reading comprehension subtests at the time of testing were considered eligible for the ERO classes. In order to be

[24]At the start of the study, it became apparent that the study schools did not have sufficient numbers of ninth-grade students in the target range to fill the ERO sections with the desired random assignment ratio. Hence, eligibility for the study was expanded to also include students reading one to two years below grade level. Sites were asked to recruit as many students as possible from the target population (two to five years below grade level) but to also recruit students whose reading skills were one to two years below grade level, so that all ERO sections could be filled in the event that a site was unable to recruit a sufficient number of students reading in the target range. The first and second ERO reports provide a more detailed discussion of the challenges related to student recruitment.

[25]Testing and recruitment began earlier in the second year of the study to ease some of the student recruitment challenges faced in the first year. In the first study year (2005-2006), districts tested ninth-grade students enrolled in the study schools at the beginning of the fall semester. In the second year of the study (2006-2007), districts began recruiting students in the spring semester, by testing all eighth-grade students enrolled in the districts' primary feeder middle schools, with the exception of one school district where students were tested in the fall of ninth grade (this district is comprised of four high schools). The second ERO report provides further details on differences between study years in terms of the timing of testing and recruitment.

part of the study, students also had to provide a signed affirmative parental consent form and a completed baseline survey.

Random Assignment of Students

From the eligible population, two cohorts of ninth-grade students were recruited into the study, one for each year of program implementation (2005-2006 and 2006-2007). Based on power calculations, the goal was to identify and recruit approximately 110 eligible and consenting students per cohort from each participating high school. Random assignment was then conducted such that, in each school, 60 students were selected to enroll in the ERO classes ("ERO group").[26] The remaining students were assigned to one of the regular elective classes available to ninth-grade students ("non-ERO group"). Although district and school staff participated in student recruitment, computerized random assignment of students was conducted solely by MDRC staff.[27]

The random assignment of students to the ERO programs makes it possible to draw valid inferences about the impact of the programs on student outcomes. Students in the non-ERO group serve 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, differences in outcomes between the ERO and the non-ERO groups represent the impact that the ERO programs had on students' outcomes over and above what these students would have achieved had they not been offered the opportunity to enroll in the reading class instead of a regularly scheduled elective class.

Characteristics of Students in the Study

As explained earlier, eligibility for the ERO classes was extended to students reading one to five years below grade level, in order to fill the class sections. However, the target population for the study consists of students reading two to five years below grade level. The findings in this report and in previous reports are based on students in this target population; all further references to the "study sample" in this report refer to students in this target range.[28]

---

[26]The ERO programs were designed to accommodate between 12 and 15 students per class; each high school was required to offer four ERO class sections, resulting in approximately 60 students served per high school.

[27]Random assignment in each of the high schools was blocked by test score range (1 to 1.9 years below grade level and 2 to 5 years below grade level).

[28]See Appendix Figure E.1 for information on the construction of the study sample (two to five years below grade level) from the pool of eligible students (one to five years below grade level).

Across both cohorts of study participants, there are 5,595 students in the study sample (on average, 82 students per school per cohort), with 57 percent randomly assigned to the ERO group and 43 percent randomly assigned to the non-ERO group.[29] On average, students in the study sample were reading four years below grade level at baseline, reflecting the program eligibility criteria, and 30 percent of them were overage for grade at the start of the study, suggesting that they had been retained in a prior school year. On average, 78 percent of students in the study sample are black or Hispanic; 47 percent of them speak a language other than English in their home; and 67 percent of them received free or reduced-price lunch. As discussed in previous study reports, random assignment resulted in a high degree of similarity between the ERO and non-ERO groups on a variety of baseline characteristics for both cohorts of study participants.[30] The characteristics of students in the full study sample will be compared with those of students in the analysis samples for this study in a later section of this chapter (see Table 2.5).

Data Sources and Measures

The ERO evaluation uses a variety of data sources to measure students' outcomes and their characteristics prior to random assignment. Information was also collected to assess the fidelity and quality of program implementation, as well as to measure the nature and quantity of literacy services offered to non-ERO students as part of "business as usual." Following is an overview of the data sources and measures utilized in this study.

Student Outcomes and Baseline Characteristics

Information on the student outcomes and baseline characteristics in the ERO study come from several sources:

- **Reading assessments:** The Group Reading Assessment and Diagnostic Examination (GRADE) assessment was used to measure students' reading achievement prior to random assignment and then again at the end of ninth

[29]There are 2,762 students in the RAAL schools in the study sample (58 percent of whom were assigned to the ERO group) and 2,833 students in the Xtreme Reading schools (57 percent of whom are in the ERO group). The study sample for Cohort 1 includes 2,916 students, while the study sample for Cohort 2 includes 2,679 students (57 percent of students were assigned to the ERO group in both cohorts).

[30]See Appendix E for tables comparing the characteristics of students in the ERO and non-ERO group in the study sample. As reported in these tables, a chi-squared test indicates that there is no systematic difference in the background characteristics of the ERO and non-ERO groups in the study sample. The lack of a systematic difference indicates that random assignment was successful in creating two equivalent research groups at baseline.

grade.[31] 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.[32] It is used widely to measure performance and track the 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.[33]

- **Student surveys:** Students in the study completed a survey at baseline and at the end of ninth grade. The baseline survey was completed by students in the study sample prior to random assignment and includes background information on such items as gender, race/ethnicity, age, and parents' education. The follow-up survey, which was administered to students at the end of ninth grade at the same time as the GRADE assessment, includes information on students' reading behaviors and attitudes.[34]

- **School records:** Participating school districts provided school records for students in the study. These records include course transcript data, scores on state standardized tests, attendance, and disciplinary records. School records data are available for the school year in which students were offered the opportunity to enroll in the ERO program ("program year") as well as the school year after program participation ("follow-up year").[35] Districts were also asked to provide records for the school year prior to the program year, which provides a source of information on students' achievement before their participation in the study. (Table 2.2 provides further detail on the years of available data for each of the two cohorts of students in the study.)

---

[31]Level H (Form A) of the reading comprehension subtests was administered to students at baseline, while Level H (Form B) was administered to students in the spring of ninth grade, in addition to the vocabulary subtest.

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

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

[34]See Appendix B for a copy of the survey instrument.

[35]School records data are also available for a second follow-up year — but for Cohort 1 only. The findings in this report focus on impacts in the program year and the first follow-up year, because school records data are available for both Cohort 1 and Cohort 2 during this time frame. Impacts on high school outcomes in the second follow-up year can be found in Appendix I (see Cohort 1 impact tables).

The Enhanced Reading Opportunities Study

Table 2.2

Data Collection Years, by Cohort

Baseline/Follow-Up Year	School Year	
	Cohort 1[a]	Cohort 2
Baseline		
School year prior to program year	2004-2005	2005-2006
Follow-up		
Program year	2005-2006	2006-2007
Follow-up year	2006-2007	2007-2008

NOTE: [a]School records data are available for a second follow-up year for Cohort 1 (2007-2008 school year). Impact findings for this additional follow-up year can be found in Appendix I.

The first two reports for this study examined the ERO programs' impact on students' reading achievement and their reading behaviors, while this report focuses on the programs' impacts on three types of high school outcomes: (1) students' performance in their core courses; (2) their performance on the high-stakes standardized tests required by their states; and (3) their school behaviors, which includes attendance and disciplinary outcomes. The remainder of this section describes these outcome measures (see also Table 2.3 for a summary). This is followed by a description of the background characteristics and prior achievement measures that are also available for describing students in the study.

Reading Achievement

A central objective of the ERO programs is to provide students with immediate and intensive instruction on the use of the strategies that expert readers use to understand written texts, which in turn should improve students' reading comprehension. The ERO programs also provide 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). Hence, the programs may also improve students' reading vocabulary.

In the previous two study reports, the impact of the ERO programs on students' reading skills was assessed using the GRADE assessment, administered at the end of the program year (ninth grade):

- **Reading Comprehension:** Students' reading comprehension at the end of the program is measured based on their average score on the two reading

The Enhanced Reading Opportunities Study

Table 2.3

Student Outcomes in the ERO Study

Outcome	Scale	Point(s) of Measurement	Source
Reading achievement[a]			
Reading comprehension score	Scaled scores	End of program year	GRADE assessment
Reading vocabulary score	Scaled scores	End of program year	GRADE assessment
Reading behaviors			
Amount of school-related reading	Prior-month occurrences	End of program year	Student survey
Amount of non-school-related reading	Prior-month occurrences	End of program year	Student survey
Use of reading strategies	4-point agreement scale	End of program year	Student survey
Course performance			
Grade point average (GPA) in core subject areas[b]	0.0 to 4.0	End of program year End of follow-up year	School records
Credits earned in core subject areas	As a percentage of core credits required for graduation in the district	End of program year End of follow-up year (cumulative)	School records
Performance on state tests			
State test scores in core subject areas[c]	Standardized (z-scores)	End of program year End of follow-up year	School records
Student behaviors			
Attendance rate	Days attended as a percentage of days enrolled	End of program year End of follow-up year	School records
Ever suspended[d]	Dichotomous (0/1) Dichotomous (0/1)	End of program year End of follow-up year	School records

(continued)

Table 2.3 (continued)

NOTES: [a]The national average for standard scores is 100, and the standard deviation is 15. The test-retest reliability of the GRADE assessment is 0.92 for ninth-grade students; the alpha reliability is 0.90-0.91 for the reading comprehension subtest and 0.80-0.86 for the vocabulary subtest as indicated in the GRADE Technical Manual (Level H).

[b]The four core subjects areas are English language arts (ELA), social studies, science, and math. In order to pool grades across school districts, students' grades in core courses were converted to a common 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0. Students' GPA in core courses was then obtained by taking the average of students' course grades in these subjects.

[c]Test scores on each assessment were standardized by district, study year, and cohort, based on the mean and standard deviation of the non-ERO group.

[d]"Ever suspended" measures the percentage of students suspended (in school or out of school) or expelled one or more times during the relevant school year.

comprehension subtests included in the GRADE — passage comprehension and sentence comprehension.

- **Reading Vocabulary:** The vocabulary subtest in the GRADE is used to assess whether the programs increase the stock of words that students know.

Reading comprehension and vocabulary scores are in standard score units provided by the American Guidance Service, which publishes the GRADE.[36]

One of the limitations of this study is that the GRADE was not re-administered at the end of the follow-up year (tenth grade). Therefore, it is not possible to directly examine whether impacts on reading comprehension in the ninth grade are sustained in the following school year. However, students' GRADE scores in the ninth grade are highly correlated with their academic performance in the tenth grade (GPA, credit accumulation, state test scores), the latter being available in this study. Therefore, the estimated impact of the programs on these latter outcomes provides some indication of whether impacts on reading comprehension persist in tenth grade.[37]

Reading Behaviors

Each of the ERO programs also endeavors to enhance students' 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 previous two reports examined the impact of the ERO programs on three measures of reading behaviors, which were constructed from responses to the student survey (administered at the end of ninth grade):

- **Amount of School-Related Reading:** The self-reported number of times during the prior month that a student read different types of text in school or for homework (7 survey items).[38]

---

[36]Specifically, 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. The test-retest reliability of the GRADE assessment (Level H) is 0.92 for ninth-grade students; the alpha reliability is 0.90-0.91 for the reading comprehension subtest and 0.80-0.86 for the vocabulary subtest (American Guidance Service, 2001b, pp. 85-87).

[37]See Appendix J for associations between the reading and the academic performance outcomes in this study.

[38]History, science, or math textbooks; literary texts; research or technical reports; newspaper or magazine articles; or workbooks.

- **Amount of Non-School-Related Reading:** The self-reported number of times during the prior month that a student read different types of text outside school (7 survey items).[39]

- **Use of Reflective Reading Strategies:** Students' reported use of the reading skills and techniques that the ERO programs try to teach as they read for their English language arts class and for one other academic class.[40] Students were asked to rate their use of these strategies on a scale from 1 (strongly disagree) to 4 (strongly agree) (4 survey items).

Further information on the survey items used to construct these three measures, as well as the reliability of the measures, can be found in Appendix B.

Performance in Core Courses

While the short-term goal of the ERO programs is to improve students' reading skills, their longer-term objective is to improve students' performance in their core courses (English language arts or ELA, social studies, science, and math). The expectation is that, by providing students with the reading strategies that they need to understand complex texts, the ERO programs will help students learn more of the subject matter taught in their core high school courses. In turn, students' course grades will improve, which will enable students to earn the credits they need to graduate.[41]

In this report, the impact of the ERO programs on course performance is assessed using two indicators, both of which are derived from students' course transcripts:

- **Grade Point Average (GPA) in Core Subject Areas:** This indicator gauges the extent to which students are learning the subject matter taught in their core courses. The measure is defined as students' average grade in

[39]Fictional books; plays; poetry; (auto)biographies; books about science, technology, or history; newspaper or magazine articles; or reference books.

[40]The 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.

[41]This report focuses on core courses — rather than elective courses — because performance in core courses is an especially important determinant of high school graduation and college readiness.

core subject areas in the program year or follow-up year and is based on a four-point scale.[42,43]

- **Credit Accumulation in Core Subject Areas:** This indicator gauges whether students' performance in their core courses is such that they are making progress toward graduation. This measure is defined as the number of core credits earned cumulatively by the end of the program year or follow-up year, as a percentage of the number of core credits required for graduation in a student's district.[44] Thus, this measure captures the extent to which students have progressed toward satisfying the graduation requirements in their district.

Performance on State Tests

GPA and credit accumulation provide an indication of whether students are learning what their teachers and school administrators expect of them. Another key question is whether the ERO programs have an impact on whether students are learning what policymakers beyond

---

[42]Course grades were converted to a common 4-point scale, in order to make it possible to pool these data across schools. The following conversion was used to convert letter grades: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0. School districts using "percentage" grades were asked to provide a method to convert these percentages to letter grades, which made it possible to apply the above conversion in these schools as well. Students' GPA in core subject areas is defined as their average numerical grade across the four core subject areas for which students have a grade (since students did not always enroll in and therefore receive a mark in all four core subjects in a given year).

[43]One of the key criticisms associated with using GPA as a measure of academic performance does not apply in the context of the ERO study. The issue is that differences in GPA between students may be partly due to differences in the grading standards of teachers and the difficulty of courses chosen by students, rather than differences in their latent academic ability. However, because the ERO study is based on random assignment of students *within schools*, course-taking patterns and grading standards should be the same on average for students in the ERO and the non-ERO groups. This means that a comparison of GPA for these two groups within schools should yield an internally valid estimate of the impact of the programs on academic performance.

The study team used school records data to examine the course-taking patterns of ERO and non-ERO students more closely (see Appendix H for detailed findings). These analyses show that, in the program year, ERO and non-ERO students attempted the same number of core credits, which confirms that their course-taking patterns were similar on average. This means that, in the program year, the ERO programs' impact on students' GPA can be interpreted as their effect on academic performance. In the follow-up year, however, ERO students attempted 0.1 *fewer* core credits than non-ERO students (effect size = -0.05; p-value = 0.032), perhaps due to the fact that they earned *more* credits in the previous year (as reported in the impact findings in Chapter 4). This means that the estimated impact of the ERO programs on GPA *in the follow-up year* is a combination of two types of impact: (1) their effect on academic performance and (2) their effect on course-taking patterns.

[44]Course catalogs and student handbooks were used to determine districts' graduation requirements.

the district expect of them, as measured by students' scores on the high-stakes standardized tests mandated by their state. The ERO programs may improve students' scores on state tests through their impact on reading comprehension and knowledge gains in the classroom. In addition, both reading interventions teach students specific reading strategies that they can use to better understand the test items on standardized assessments.

The impact of the ERO programs on students' performance on state tests is assessed in this report using the following indicator:

- **Scores on State Tests in Core Subject Areas:** This indicator gauges the extent to which students demonstrate proficiency in the content areas deemed important by state policymakers.[45,46] In order to make it possible to pool data from different state assessments, scores for each assessment are standardized (z-scored) by district, by year (program year or follow-up year), and by cohort, based on the mean and standard deviation of the non-ERO group.[47]

Student Behaviors

While the primary focus of the reading programs is on the teaching of reading strategies, both interventions also strive to instill positive behavioral norms. By this token, the ERO programs may also improve students' behavior in the classroom and at school. In this report, the impact of the ERO programs on student behaviors is assessed using the following indicators:

- **Attendance Rate:** This indicator is defined as the number of days that a student attends school during the program year or follow-up year, as a percentage of the days that the student is enrolled during that year.

- **Ever Suspended:** This indicator gauges whether students exhibit school or classroom behaviors that are detrimental to their own learning or to that of other students. It is a dichotomous indicator of whether a student was ever

[45]This includes assessments administered for school accountability purposes (No Child Left Behind), as well as tests that are high-stakes for the student (graduation tests and end-of-course tests). It also includes general achievement tests (for example, the TAKS science exam in Texas) as well as tests that are course-specific (such as the Earth Sciences Regents exam in New York State). Appendix D provides further information on the assessments used by states in this study.

[46]If a student took a given state test more than once (for example, because he or she failed on the first attempt), only his or her first score is used in the analysis.

[47]Appendix D provides further information on the standardization of test scores.

suspended (either in-school or out-of-school) during the program year or the follow-up year. [48,49]

Baseline Characteristics and Prior Achievement

Information was also collected on students' background characteristics and prior achievement, for the purposes of describing the sample and to use as covariates in the impact model:

- **Reading comprehension at baseline:** As noted earlier, the GRADE assessment was administered to all students prior to random assignment in order to identify eligible students. Two reading comprehension subtests were administered — passage comprehension and sentence comprehension. Students' baseline reading comprehension scores provide a measure of their reading comprehension skills before the start of the program.

- **Background characteristics:** All students in the study also completed a baseline survey prior to random assignment. The baseline survey includes information on students' background characteristics, such as gender, race/ethnicity, age, and parents' education.

- **Prior achievement:** As part of the school records request, districts in the study were asked to provide students' records for the school year immediately prior to their participation in the study, when most students were in the eighth grade.[50] This provides a source of information on students' academic outcomes (course performance, state test scores) and attendance before they were recruited into the study. This information was not yet available at the time of the first two study reports.

---

[48]More accurately, "ever suspended" is a dichotomous indicator of whether students were suspended *or expelled* during the school year. Given that most students who were expelled during the year were also suspended at some point during that year, this outcome is characterized as "suspensions" in this report.

[49]One school district (four schools) provided information on the disciplinary infractions committed by students but not the disciplinary actions taken by the school. In this district, a student is considered to have been "suspended" if they committed five or more infractions of any type during the school year or if they committed at least one serious infraction (including fighting, harassment and bullying, insubordination and persistent disobedience, physical violence or aggression, reckless endangerment, refusing to stay in detention, rioting, threats, truancy, vandalism, and possession of weapons).

[50]For Cohort 2 students in one school district (four schools), school records are from *two* years prior to the program year (2004-2005 school year) when most of these students were in the seventh grade because these data are more consistently available.

Implementation Fidelity and Service Contrast

The first two study reports presented findings on the fidelity with which the sites participating in the ERO study implemented the models of the two supplemental literacy programs. In addition, these reports also included an assessment of the service contrast in this evaluation — that is, the extent to which the literacy instruction received by students in the ERO group was substantially different than that of the students in the non-ERO group. Following is a brief description of the data sources used to assess implementation fidelity and the service contrast.

Implementation Fidelity

The analysis of ERO program implementation fidelity is based on field research visits to each of the 34 high schools in the spring of the first implementation year (spring 2006) and the fall and spring of the second implementation year (2006-2007 school year). The primary data collection instrument for the site visits was a set of protocols for classroom observations.[51] The observation protocols provided a structured process for trained classroom observers to rate the learning environments and reading instruction in the ERO classrooms. These constructs included in the protocol 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.[52] For each site visit, ratings on the constructs in the protocol were combined to calculate composite measures for two key dimensions of implementation: (1) learning environment and (2) comprehension instruction. These summary measures will be described in greater detail in the next chapter, where the key implementation findings from the two previous reports are discussed. Appendix C provides further background on the properties of the classroom observation data and the fidelity measures.

Dosage and Service Contrast

The impact of the ERO program is hypothesized to be a function of the amount of ERO instruction received by students in the ERO group (that is, dosage), as well as the contrast between the ERO and non-ERO groups with respect to the amount of other literacy supports received. Several types of data were collected and analyzed to measure dosage and service contrast in this study:

---

[51] These observation protocols can be found in Appendix C.

[52] Biancarosa and Snow (2004).

- **ERO class attendance records:** Each of the ERO teachers provided ERO class attendance data for students assigned to an ERO class. These data were used to determine the average number of hours of ERO instruction received by students in the ERO group (a measure of "dosage").

- **Student course schedules:** Each school provided the study team with copies of the schedules for all students in the study sample. These data were used to confirm that ERO students were enrolled in the ERO classes and that non-ERO students were not. These data also made it possible to compare the types of core and non-ERO elective classes taken by students in the ERO and non-ERO groups.

- **Student survey:** The student survey administered at the end of ninth grade was used to measure students' participation in the following supplementary literacy support activities: (1) a school-based literacy class, (2) tutoring from an adult at the school, (3) a literacy class outside of school, and (4) tutoring from an adult outside of school. The purpose of these four measures is to determine whether non-ERO students received some form of supplemental literacy support, and the extent to which such supports may have reduced the service contrast between the ERO and non-ERO groups. The construction of these measures is described in greater detail in Appendix B.

Analysis Samples in the Study

In this study, impacts on student outcomes are estimated based on the largest sample of students for whom data are available on the outcome of interest. This section discusses the availability of school data for each type of outcome in the study and describes the samples of students ultimately used to estimate impacts in this report and in previous reports.

Availability of Data

Table 2.4 presents information on the availability of data in the program year and follow-up year, by data source and type of outcome.[53] There are several important patterns to note from this table.

[53]Appendix E presents tables showing the availability of data for schools implementing RAAL and Xtreme Reading, separately. The response patterns for these two subgroups of schools are similar to those discussed in this section.

The Enhanced Reading Opportunities Study

Table 2.4

Availability of Data for the Full Study Sample, by Data Source and Type of Outcome

Measure	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
GRADE reading assessment				
Program year (%)	83.4	80.2	3.2*	0.002
Student survey				
Program year (%)	83.3	79.8	3.6*	0.001
School records				
GPA and credit accumulation				
Program year (%)	91.7	92.5	-0.9	0.231
Follow-up year (%)	79.3	79.2	0.1	0.927
Performance on state tests				
Program year (%)				
English language arts	39.8	39.5	0.3	0.693
Social studies	17.5	17.3	0.2	0.675
Science	41.6	42.6	-1.0	0.337
Math	47.9	47.1	0.8	0.427
Follow-up year (%)				
English language arts	43.5	42.7	0.7	0.456
Social studies	40.9	40.1	0.8	0.461
Science	48.3	47.3	1.1	0.384
Math	46.0	44.8	1.2	0.297
School behaviors				
Program year (%)				
Attendance rate	90.7	91.7	-1.0	0.177
Ever suspended	81.9	82.2	-0.3	0.599
Follow-up year (%)				
Attendance rate	78.5	78.8	-0.2	0.838
Ever suspended	70.4	70.0	0.4	0.682
Sample size[a]	3,204	2,391		

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

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]The sample sizes reported here are for the full study sample.

First, students in the study sample are most likely to have information on their performance in core courses (GPA and credit accumulation). This is due to the fact that if a student is enrolled at the school for at least one semester during the school year, the school district will have course transcript information for that student.[54] That said, the availability of course records data decreases from the program year to the follow-up year, due to student mobility across school districts, as well as students dropping out of school.[55] On average across both program groups, data on course performance are available for 92 percent of the study sample in the program year; 79 percent of students in the study sample also have course records data in the follow-up year.

Second, Table 2.4 shows that information on students' behavioral outcomes (attendance and suspensions) is almost as likely to be available as course performance data. In the program year, for example, 91 percent of students in the full study sample have attendance data (compared with the 92 percent who have course performance data). The availability of information on suspensions is somewhat lower. In the program year, information on this measure is available for 82 percent of students in the full study sample. This is due to the fact that disciplinary data are not available for five schools in the study.

Third, Table 2.4 shows that among the different types of student outcomes, students are least likely to have a state test score for a particular subject area in a given year. In fact, there is no subject area or grade for which state test scores are available for all of the study districts. As a result, less than 50 percent of students in the full study sample have state test data in any given year. In the program year, for example, the percentage of students in the study sample for whom state test data are also available ranges from 17 percent to 48 percent, depending on the subject area.

These patterns are explained by differences in the testing policies of the states that are represented in the ERO study. In the first instance, there is variation across states in terms of the grade in which a given subject area is tested. For example, in some states, a state ELA test may be administered in the ninth grade, while in other states a test for this subject may not be administered until the tenth or eleventh grade. A second important feature of states' testing policies is that some of them administer "end-of-course" tests, in which case students take the state test in a given subject area once they have completed the associated course. This means, for example, that some students will take the state's biology exam in ninth grade while others

[54]Because school records are provided by school districts, data are also available for students in the study sample who transferred to another school in the district during the study period.

[55]The information provided by districts does not make it possible to determine what percentage of students fall into each of these two categories.

will take it in tenth grade, depending on when they decide to enroll in biology. Hence, within school districts, there is variation between students in terms of when they take the state test.[56]

Finally, Table 2.4 shows that in the full study sample, students in the ERO group completed the GRADE reading assessment and student survey (administered at the end of the program year) at a higher rate than students in the non-ERO group. Specifically, testing and survey data are available for 83 percent of students in the ERO group, while these data are available for 80 percent of students in the non-ERO group; this difference of 3 percentage points is statistically significant.

Students in the Analysis Samples

As noted earlier, impacts in this study are estimated based on the largest sample of students for whom data are available on the outcome of interest. This means that:

- **The analysis sample varies across the different types of outcome in this study.** In the two prior study reports, impacts on reading achievement were based on students who completed the GRADE test at the end of ninth grade, while impacts on reading behaviors were estimated using students who completed the student survey at the end of ninth grade. In this report, impacts on course performance are estimated based on students with data on GPA and credit accumulation, while impacts on state test scores and school behaviors are based on the subset of students with course performance data who also have information on state tests or school behaviors.

- **For impacts on high school outcomes (course performance, state tests, and student behaviors), the analysis sample differs from the program year to the follow-up year.** That is, impacts in the program year are estimated based on students with school records in that year, while impacts in the follow-up year are based on the subset of the students who also have school records in the follow-up year.

Because the analysis sample depends on the outcome of interest, the remainder of this

[56]Among students with state test scores, there is no statistically significant difference between the ERO and non-ERO group with respect to the grade or year in which they take the state test for a given subject area.

report will refer to the following analysis samples:[57]

- **GRADE respondent sample:** This analysis sample includes students who completed the GRADE reading assessment at the end of the program year. It includes 4,584 students across both cohorts of study participants (representing 82 percent of the full study sample). As noted earlier, the percentage of students who have a GRADE test score differs by a statistically significant amount between the ERO and the non-ERO group.

- **School records sample:** This analysis sample includes students who, at minimum, have course performance data. In the program year, the school records sample includes 5,150 students (representing 92 percent of the full study sample); 4,436 of these students also have course data in the follow-up year (representing 79 percent of the full study sample).[58] The percentage of students for whom school records data are available does not differ by a statistically significant amount between the ERO and the non-ERO group, whether in the program year or the follow-up year.[59]

Given that impacts are based on students for whom outcome data is available, it is important to examine two key questions. First, do students in the analysis samples differ from students in the entire full study sample and from students excluded from the analysis samples? The answer to this question determines whether the findings presented in this report can be generalized to the full study sample. Second, within the analysis samples, are the ERO group and the non-ERO group still equivalent? The answer to this question affects whether the impact

---

[57]Another analysis sample in this study — but which is not listed here — is the "survey respondent sample" used to estimate impacts on reading behaviors. Because the GRADE assessment and survey were administered at the same point in time (end of ninth grade), there is a high degree of overlap between the GRADE respondent sample and the survey respondent sample. Specifically, the survey respondent sample includes 4,573 students, while the GRADE respondent sample includes 4,584 students; 4,557 students completed both the GRADE test and survey.

[58]In RAAL schools, the school records sample includes 2,563 students in the program year (representing 93 percent of the full study sample in those schools) and 2,212 students in the follow-up year (representing 80 percent of the full study sample in those schools). In Xtreme Reading schools, the school records sample includes 2,587 students in the program year (representing 91 percent of the full study sample in those schools) and 2,224 students in the follow-up year (representing 79 percent of the full study sample in those schools).

[59]Appendix H presents impacts in the program year for students who are in the school records sample in *both the program year and the follow-up year* (longitudinal sample). These analyses were conducted to make sure that the pattern of impacts presented in the next chapter — for the program year and the follow-up year — are not driven by differences in sample size or sample composition across the two years. These supplemental analyses show that the pattern of findings based on the longitudinal sample is the same as the pattern of findings for the school records sample.

findings in this study represent the effect of the ERO programs and not the effect of other unobserved factors.

Comparison of the Analysis Samples with the Full Study Sample

Table 2.5 compares the background characteristics and prior achievement of students in the analysis samples and the full study sample. As seen in this table, the school records sample and the GRADE respondent sample reflect the characteristics of students in the full study sample. The typical student in the analysis or study samples, for example, was reading well below grade level at baseline, with a reading comprehension score of 85 standard score points, corresponding to a 5.0 grade level (four years below grade level) and to the fifteenth to sixteenth percentile nationally. In the school year prior to participation in the study (the eighth grade for most students), the average student earned 87 to 89 percent of core credits attempted (depending on the sample) and had a GPA in the range of 2.0 to 2.1, corresponding to a C average. On average, students in the samples also had several characteristics associated with the risk of doing poorly in school. On average, 47 to 48 percent reported that a language other than English is spoken in their homes, and 67 to 69 percent were eligible for free or reduced-price lunch. In addition, 77 to 78 percent of students are Hispanic or black. Given the similarity between the full study sample and the analysis samples — and the fact that almost all students in the full study sample are included in the analysis samples — estimated impacts in this report are likely to be generalizable to the full study sample.[60]

That said, caution should be exercised when generalizing estimated impacts on state test scores to the full study sample. As noted earlier, in any given follow-up year, only a subset of the study districts administers state tests; moreover, in most districts students only take the test when they have completed the relevant coursework, such that less than 50 percent of students in the full study sample have state test data in any given year. As a result, the characteristics of students for whom state test data are available does not necessarily reflect the characteristics of the full study sample.[61] For example, the percentage of students who come from multilingual

---

[60]However, one should be cautious when generalizing the impact findings in this report to students *without* school records or GRADE test data, since there are notable differences between students with and without this information. In particular, students without data on a particular outcome are more likely to be doing poorly in school. These students had lower reading comprehension test scores at baseline and are more likely to be overage for the ninth grade. Additionally, these students had lower grade point averages in core subjects, earned fewer core credits, and attended school less in the school year prior to ERO. See Appendix E for a detailed comparison of students with and without outcome data. Results are presented for all the participating high schools together and separately for the groups of schools using RAAL and Xtreme Reading.

[61]See Appendix E for a comparison of students in the state test samples with students in the full study sample.

The Enhanced Reading Opportunities Study

Table 2.5

Characteristics of Students in the Analysis Samples Relative to Students in the Full Study Sample

Characteristic	Full Study Sample	GRADE Respondent Sample	School Records Samples: Program Year Sample	School Records Samples: Follow-Up Year Sample
Baseline student survey and testing[a]				
Race/ethnicity (%)				
Hispanic	31.0	32.5	31.9	31.3
Black, non-Hispanic	46.8	44.7	46.4	46.3
White, non-Hispanic	16.0	16.5	15.4	15.9
Other	6.2	6.3	6.3	6.4
Male (%)	50.5	51.0	50.4	50.0
Average age (years)	14.8	14.7	14.8	14.7
Overage for grade[b] (%)	29.9	26.4	28.8	25.8
Language other than English spoken at home (%)	47.1	48.3	47.6	47.3
Mother's education level (%)				
Did not finish high school	19.2	18.1	19.0	18.1
High school diploma or GED certificate	26.0	26.2	25.7	26.5
Completed some postsecondary education	32.1	32.6	32.3	32.3
Don't know	22.6	23.2	23.0	23.1
GRADE reading comprehension[c]				
Average standard score	85.2	85.4	85.3	85.4
Corresponding grade equivalent	5.0	5.1	5.0	5.1
Corresponding percentile	15	16	15	16
School records prior to program year[d]				
GPA in core subjects (out of 4.0)				
English language arts (ELA)	2.02	2.08	2.04	2.07
Math	1.87	1.94	1.90	1.93
Social studies	2.04	2.11	2.06	2.09
Science	1.94	2.01	1.96	1.99
Credits earned in core subjects (as a percentage of credits attempted)	87.2	89.1	87.7	88.6
Attendance rate (%)	94.6	95.4	94.7	95.2
Free and reduced-price lunch (%)	67.1	68.3	68.3	68.5
Sample size	5,595	4,584	5,150	4,436

(continued)

Table 2.5 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The values in these columns are the observed means for students in the study sample and the analysis samples. The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]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 A).

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full study sample and school records sample.

homes ranges from 44 to 60 percent in the state test sample (depending on subject area and the follow-up year) compared with 47 percent for the full study sample. This means that the estimated impact of the ERO programs on state test performance in this report may not generalize to all students in the study sample. Caution should also be exercised when comparing estimated impacts on state test performance across subject areas and years, since each impact finding for state tests is based on a different subset of schools.

Comparison of the ERO and non-ERO Groups in the Analysis Samples

Next the research team examined whether, in each of the analysis samples in this study, students in the ERO group differ systematically from students in the non-ERO group in terms of their background characteristics and prior achievement. As noted earlier in this section, the percentage of students who completed the GRADE reading follow-up test differs by a statistically significant amount between the ERO and the non-ERO groups. Thus, an important concern is whether, in the GRADE respondent sample, students in the non-ERO group still provide the right "counterfactual" for students in the ERO group.

However, baseline analyses indicate that there is a high degree of similarity between the ERO and the non-ERO groups in all of the analysis samples used in this study, including the GRADE respondent sample.[62] Thus, the analysis samples preserve the balance that was achieved with random assignment for the full study sample, such that differences in student outcomes between the two groups reflect the impact of the ERO programs rather than preexisting differences in students' background characteristics and/or prior achievement. This is also true for each of the groups of schools using the two supplemental literacy programs (RAAL and Xtreme Reading).

Analytic Methods and Procedures

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 outcomes. It first briefly describes the statistical model used to estimate impacts in this study and then 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 concludes by discussing the approach taken in this report with regards to multiple hypothesis testing (that is, the strategy used to avoid concluding that an estimated impact is statistically significant when, in fact, the true impact is zero).

Statistical Model for Estimating Impacts

Because students were randomly assigned to the ERO class, the impact of the ERO programs can be estimated by comparing the outcomes of students in the ERO and the non-ERO groups. In practice, ordinary least squares regression is used to estimate the difference in outcomes between the ERO and the non-ERO groups, which makes it possible to account for the blocking of random assignment by school and cohort.[63] Following are some of the key features of the statistical model used to estimate impacts in this report and in the two previous reports (see Appendix F for further details):

- **Covariates.** In order to improve the precision of the impact estimates, the analysis controls for random differences between the ERO and the non-ERO

[62]Omnibus tests indicate that, for each of the analysis samples, the ERO and the non-ERO groups are not systematically different in terms of their baseline characteristics. See Appendix E for details.

[63]Controlling for random assignment blocks in the model also accounts for the clustering of student outcomes by school and cohort because it explains all of the between-school and between-cohort variation in student outcomes.

groups in terms of the following covariates:[64] students' GRADE reading comprehension test score at baseline, whether students were overage for grade at the start of ninth grade (and likely to have been retained in a prior grade), students' standardized test scores in reading and math in the school year prior to participation in the study, and a measurement of the outcome of interest in the school year prior to study participation.[65] The choice of these covariates was made prior to estimating impacts, based on the fact that they are highly predictive of high school outcomes.[66]

- **Block fixed-effects.** The impact model treats random assignment blocks as fixed effects. This means that the findings presented in this report should be interpreted as the estimated impact of the ERO programs in the set of schools in which these interventions were implemented. In other words, the impact estimates are not generalizable to other schools or school districts, nor to alternate assignments of the two interventions (RAAL, Xtreme Reading) among the schools in the study.

The impact analyses conducted for this study yield intent-to-treat estimates of the ERO programs. Some students assigned to the ERO programs chose not to attend any ERO classes (4 percent of students in the first cohort and 9 percent in the second cohort, see Chapter 3). Thus, the findings in this report represent the estimated impact of offering students the opportunity to enroll in the reading classes ("intent to treat"), rather than the impact of the reading programs on students who actually enrolled ("treatment on the treated"). However, because students'

[64]In theory, it is not strictly necessary to control for these baseline characteristics because random assignment should ensure that students in the ERO and non-ERO groups are similar in expectation at baseline. However, controlling for baseline characteristics can improve the *precision* of the impact estimates. As a sensitivity test, the statistical model was also specified without controlling for students' background characteristics and prior achievement. See Appendix F for these "unadjusted" impact estimates.

[65]For example, when estimating impacts on GPA in core subject areas, the model controls for GPA in core subject areas in the year before study participation (eighth grade for most students). Note that prior standardized test scores and a prior measurement of the outcomes were not included as covariates when estimating impacts on reading achievement and reading behaviors in prior study reports because these baseline data were not available at the time of analysis.

[66]The model also includes indicators of missing information for the baseline covariates created from school records data (state test scores and the prior measurement of the outcome). Information on background characteristics (student survey) and reading comprehension at baseline (GRADE assessment) are available for all students because these pieces of information were a requirement for participating in the study.

participation in educational interventions is typically voluntary, intent-to-treat estimates of the impact of offering a program or service are also policy-relevant.[67]

Throughout this report, impact estimates are presented both in their original metric and as an effect size. The "effect size" is a metric that is widely used for measuring the impact of education programs. It is defined as the impact (effect) of a program divided by the standard deviation of the outcome of interest. For example, an effect size of 0.20 represents an improvement in student outcomes that is equal to 20 percent of the standard deviation of the student-level distribution for that particular outcome. In this report, effect sizes are calculated by dividing the estimated impact on a given outcome (in the program year or follow-up year) by the standard deviation of that outcome for students in the non-ERO group in the analysis sample (in the program year or follow-up year).[68] The standard deviation for the non-ERO group reflects the expected variability in the outcome 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 achievement or behaviors.[69] (See also Box 2.1 for a detailed explanation of how outcome levels for the ERO and non-ERO groups are presented throughout this report.)[70]

As explained in Chapter 1, the primary purpose of the ERO study is to evaluate the impact of full-year supplemental reading programs targeted at struggling adolescent students. Thus, the main findings in this report focus on the combined of impact of RAAL and Xtreme Reading, both of which are part of the same broad class of literacy intervention. Impact findings are also presented for each reading program separately, in order to provide context for interpreting the combined impact of the two programs. Tests of the difference in impacts between the programs are used to verify that their impacts do not differ by a statistically significant amount and that it is appropriate to pool the two programs together. It is important to note that while

[67]The estimated effect of the "treatment on the treated" can be obtained by dividing the "intent to treat" impact estimates in this report by the percentage of students in the ERO group who actually attended the program (96 percent in Cohort 1 and 91 percent in Cohort 2). See Bloom (2006) for a discussion.

[68]Note that for subgroup-specific impacts (such as impacts for Xtreme Reading or RAAL schools separately), the standard deviation of the non-ERO group *for all schools and students in the school records sample pooled together* is used to calculate the effect size. Appendix F provides a full list of the non-ERO standard deviations used to calculate effect sizes in this report. It also includes standard deviations for the ERO group, and for the ERO and non-ERO groups pooled together, for use in future meta-analyses.

[69]State tests scores are standardized (z-scored) to make it possible to pool data across districts. Hence, impacts on state test scores are presented only in effect size units.

[70]Confidence intervals and standard errors for the key impact findings can be found in Appendix F.

Box 2.1

Description of the Calculation and Presentation of Outcome Levels

Throughout the report, when a table is presented to report estimated program impacts, the mean outcome levels for the ERO and the non-ERO groups are reported, in order to provide context for interpreting the estimated differences. Program impacts are estimated using a regression model that uses all available observations from both the ERO group and the non-ERO group, and the mean outcome levels are calculated using the same impact regression model.

When calculating the regression-adjusted mean outcome levels for the ERO and non-ERO groups, the adjustment is made using the observed mean covariate values for the ERO group in the impact regression model. In other words, means for *both* groups are "regression-adjusted" using this common set of baseline covariate values: the *ERO group's observed means.*

By adjusting based on the observed mean covariate values for the ERO group, the tables report:

1. Observed mean outcome levels for students randomly assigned to the ERO group
2. Regression-adjusted mean outcome levels for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment

By presenting the observed mean outcome values for the ERO group, the discussion is based on the actual mean outcomes for the ERO group, which makes it possible to compare these actual values with those for other reference groups or for the same group of students over time. The reported mean outcome level for the non-ERO group also has a straightforward interpretation: it provides an unbiased estimate of how the ERO group students would have performed had they not been assigned to the ERO programs. In other words, it represents the "counterfactual."

In the text and tables of this report, when presenting these outcome levels, the *observed* mean level for the ERO group is referred to as the "ERO group" mean. The mean value for the counterfactual, or the *regression-adjusted* mean for the non-ERO group, is referred to as the "non-ERO group" mean. In addition, note that observed means (adjusted only for random assignment blocks) for both the ERO and the non-ERO group are presented in Appendix Tables F.2 and F.3.

these statistical tests are appropriate as "pooling tests," they should not be used to make inferences about the true differential impact between the two programs.[71]

The new impact findings presented in this report are also pooled across the two participating cohorts of students. As found in previous reports, the estimated impact of the ERO programs on students' reading comprehension scores does not differ by a statistically significant amount across the two cohorts (effect size = 0.09 for Cohort 1 and 0.08 for Cohort 2).[72]

Sample Sizes and Statistical Power

An important goal for the design of the ERO study was to ensure that sample sizes would be sufficient to be able to detect program effects (if they exist) that are meaningful in students' lives and relevant to policy debates about the efficacy of supplemental literacy interventions.[73]

A common way to convey a study's statistical power is through the "minimum detectable effect" (MDE) or the "minimum detectable effect size" (MDES). Formally, the MDE is the smallest true program impact that can be detected with a reasonable degree of power (in this case, 80 percent) for a given level of statistical significance (in this case, 5 percent for a two-tailed test). The MDES is the minimum detectable effect scaled as an effect size — or, in other words, it is the MDE divided by the standard deviation of the outcome of interest. The number of students is a crucial factor that determines the degree to which the impacts on student outcomes can be estimated with enough precision to reject with confidence the hy-

---

[71]This is due to the fact that a "fixed-effects" model was chosen for the impact analysis. In this type of model, the assignment of schools to programs (RAAL, Xtreme Reading) is fixed, which means that the school-level randomization feature of the study design is essentially removed from the estimation of impacts. Had the purpose of the study been to test for a *differential* impact, the analysis would have had to allow for variability in the assignment of programs to schools (that is, a school random-effects model would have been used).

[72]Appendix I provides cohort-specific results for readers interested in these findings.

[73]There are no universally agreed-upon standards for what constitutes "small" versus "large" impacts. A meta-analysis of treatment effectiveness studies sheds some light on this issue (Lipsey, 1990). 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 0.55; and the top third of impacts ranged from 0.56 to 1.20. Under these "rules of thumb," an MDES of 0.32 would be considered small. More recent work by Hill, Bloom, Black, and Lipsey (2008) 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 0.27 effect size for a full year of school. 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.

pothesis that the program had no effect. In general, larger sample sizes provide more precise impact estimates.[74]

Table 2.6 presents the MDE and the MDES for the reading achievement and reading behaviors outcomes in this study, while Table 2.7 presents this information for the course performance and school behaviors outcomes that are the focus of this report. Table 2.8 presents the MDES for state test outcomes.[75] The minimum detectable effects in these tables are based on the number of students in the analysis samples across both cohorts (as opposed to the initial assumptions that guided the study design) and the standard errors of impact estimates. Hence, the values in these tables represent the actual precision of the analyses in this report,[76] as seen in these tables:

- **Reading achievement:** Across both cohorts, the study is equipped to detect impacts (in effect size) on reading achievement as small as 0.08. For each of the two supplemental reading programs separately (RAAL and Xtreme Reading), the study can detect impacts with effect sizes as small as 0.11.

- **Grade point average:** The study is equipped to detect impacts on GPA in core subject areas as small as 0.06 point (out of a maximum of 4.0 points) in the program year and 0.07 in the follow-up year. For each of the two reading programs separately, it can detect impacts of 0.08 point in the program year and 0.09 point in the follow-up year.

- **Credit accumulation:** The study is equipped to detect impacts on core credits earned (as a percentage of core credits required for graduation) as small as 0.7 percentage point in the program year and 1.2 percentage points in the follow-up year. For each of the two reading programs separately, it can detect impacts of 0.9 percentage point in the program year and 1.7 percentage points in the follow-up year.

---

[74]Appendix G provides a more detailed assessment of the statistical power of the ERO study's impact analysis, both in the design phase and its actual power. It also provides estimates — based on the ERO study data — of the key parameters that determine a study's MDES (that is, the explanatory power of the covariates and intraclass correlations), which may be useful for the planning and design of future studies.

[75]The MDE is not presented for state test scores because, by definition, these test scores are scaled as effect size units.

[76]Note that, even though the MDE for a given outcome is *x*, an estimated impact smaller than *x* can still be found to be statistically significant. This is because the calculation of the MDE incorporates not only the probability of making a Type I error (that is, concluding that there is an impact when in fact there is not) but also the probability of making a Type II error (that is, concluding that there is no impact when in fact the program was effective).

The Enhanced Reading Opportunities Study

Table 2.6

Minimum Detectable Effect (MDE) and Effect Size (MDES) for Impacts on Reading Achievement and Reading Behaviors, Follow-Up Respondent Samples (GRADE and Student Survey)

	Pooled Cohorts		
	Sample Size	MDE	MDES
All schools			
GRADE reading assessment			
Reading comprehension (standard score)	4,584	0.79	0.08
Reading vocabulary (standard score)	4,584	0.79	0.08
Reading behaviors			
Amount of school-related reading (prior-month occurrences)	4,484	3.40	0.08
Amount of non-school-related reading (prior-month occurrences)	4,535	2.65	0.08
Use of reading strategies (4-point scale)	4,463	0.05	0.08
Reading Apprenticeship schools			
GRADE reading assessment			
Reading comprehension (standard score)	2,255	1.13	0.11
Reading vocabulary (standard score)	2,255	1.10	0.11
Reading behaviors			
Amount of school-related reading (prior-month occurrences)	2,206	5.02	0.11
Amount of non-school-related reading (prior-month occurrences)	2,233	3.84	0.11
Use of reading strategies (4-point scale)	2,193	0.07	0.12
Xtreme Reading schools			
GRADE reading assessment			
Reading comprehension (standard score)	2,329	1.12	0.11
Reading vocabulary (standard score)	2,329	1.14	0.11
Reading behaviors			
Amount of school-related reading (prior-month occurrences)	2,278	4.61	0.12
Amount of non-school-related reading (prior-month occurrences)	2,302	3.67	0.12
Use of reading strategies (4-point scale)	2,270	0.07	0.12

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

NOTE: The MDE and MDES in this table are calculated from the relevant sample size and the standard error of the impact estimate (see Appendix G for details). Values are for both cohorts of participating students pooled together (Cohort 1 and Cohort 2).

The Enhanced Reading Opportunities Study

Table 2.7

Minimum Detectable Effect (MDE) and Effect Size (MDES) for Impacts on Course Performance and School Behaviors Outcomes, School Records Sample

	Program Year			Follow-Up Year		
	Sample Size	MDE	MDES	Sample Size	MDE	MDES
All schools						
Course performance						
GPA in core subject areas (4-point scale)	5,150	0.06	0.06	4,436	0.07	0.07
Credits earned in core subject areas (%)	5,150	0.66	0.07	4,436	1.20	0.07
School behaviors						
Attendance rate (%)	5,101	0.63	0.06	4,348	1.09	0.08
Ever suspended (%)	4,597	3.56	0.08	3,936	3.87	0.08
Reading Apprenticeship schools						
Course performance						
GPA in core subject areas (4-point scale)	2,563	0.08	0.08	2,212	0.09	0.10
Credits earned in core subject areas (%)	2,563	0.94	0.09	2,212	1.73	0.10
School behaviors						
Attendance rate (%)	2,528	1.00	0.09	2,160	1.53	0.11
Ever suspended (%)	2,346	4.88	0.10	2,016	5.33	0.11
Xtreme Reading schools						
Course performance						
GPA in core subject areas (4-point scale)	2,587	0.08	0.08	2,224	0.09	0.10
Credits earned in core subject areas (%)	2,587	0.92	0.10	2,224	1.67	0.10
School behaviors						
Attendance rate (%)	2,573	0.75	0.09	2,188	1.56	0.11
Ever suspended (%)	2,251	5.20	0.11	1,920	5.64	0.12

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

NOTES: The MDES and MDE in this table are calculated from the relevant sample size and the standard error of the impact estimate (see Appendix G for details).

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The cumulative number of credits earned is scaled as a percentage of the total number of core credits required for graduation in a student's district.

The Enhanced Reading Opportunities Study

Table 2.8

Minimum Detectable Effect Size (MDES) for Impacts on State Test Performance, School Records Sample

Outcome	Program Year		Follow-Up Year	
	Sample Size	MDES	Sample Size	MDES
All schools				
English language arts (ELA)	2,244	0.10	2,408	0.10
Social studies	952	0.18	2,237	0.11
Science	2,348	0.12	2,661	0.10
Math	2,668	0.09	2,537	0.11
Reading Apprenticeship schools				
English language arts (ELA)	1,053	0.15	1,140	0.16
Social studies	447	0.26	1,089	0.16
Science	1,151	0.19	1,269	0.14
Math	1,263	0.14	1,226	0.16
Xtreme Reading schools				
English language arts (ELA)	1,191	0.13	1,268	0.14
Social studies	505	0.24	1,148	0.15
Science	1,197	0.15	1,392	0.16
Math	1,405	0.12	1,311	0.16

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

NOTE: The MDES and MDE in this table are calculated from the relevant sample size and the standard error of the impact estimate (see Appendix G for details).

- **State test scores:** Even though state test scores are less consistently available than other student outcomes, the study is still equipped to detect impacts on state test scores (in effect size) ranging from 0.09 to 0.18 in the program year and 0.10 to 0.11 in the follow-up year, depending on the subject area. The samples for each of the two reading programs can detect impacts ranging from 0.12 to 0.26 in the program year and 0.14 to 0.16 in the follow-up year.

Hence, the analysis samples in this study are sufficiently large to fulfill the objectives of the evaluation, which is to estimate the joint impact of RAAL and Xtreme Reading on student outcomes, as well as the impact of these two programs separately.[77] As explained in other parts of this report, the purpose of the ERO study is not to test for a differential impact between the two programs, so the study's design only allows for the detection of relatively large differences in impacts between the two programs.[78]

Statistical Significance and Multiple Hypothesis Testing

The statistical significance of the impact estimates in this report and in the two prior reports is assessed using a two-tailed t-test. Statistical significance is a measure of the degree of certainty that one may have that a program's impact is actually nonzero. 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 nonzero estimate is more likely to be a product of chance. In 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, however, it is important to recognize the potential problems associated with conducting multiple hypothesis tests. Specifically, when hypothesis tests are conducted for estimated impacts on several different outcomes and for many subgroups of students and schools, this increases the likelihood of concluding that a given impact estimate is statistically significant, when in fact the program has no impact on student outcomes (this is known as a "Type I" error or a "false positive"). While it is important to avoid making conclusions based on such errors, the analysis should also not be so conservative with respect to producing false positive results that it unduly increases the likelihood of missing true impacts when they exist (that is, relying on "false negative" results or "Type II" error).

---

[77]In the planning phase for the study, target sample sizes were chosen to enable the study to detect an impact (in effect size) for each cohort as small as 0.06 for both programs together and 0.09 for each program separately. An effect size of 0.06 is equivalent to an impact of about 3 percentage points on a dichotomous outcome, such as being promoted to the next grade (assuming a maximum standard deviation of 0.5), and to an impact of 2 Normal Curve Equivalent (NCE) points on a nationally norm-referenced achievement test. Differences between the actual MDES and the "planned" MDES arise because the latter calculations were based on assumptions about the samples size, random assignment ratio, and the predictive power of pretest scores.

[78]For example, the study can detect a difference in impacts (in effect size) of the following magnitude in the program year: 0.15 for reading comprehension, 0.12 for GPA in core subject areas, and 0.13 for core credit accumulation.

When evaluating the effect of the ERO programs on high school outcomes (Chapter 4), two sets of safeguards are used to attenuate the risk of drawing inappropriate conclusions about program effectiveness.[79] The first safeguard is to confine the analysis to a parsimonious list of high school outcomes and to identify a set of "primary" outcomes and subgroups before beginning the impact analysis. All other outcomes and subgroups are considered "secondary" and used either to contextualize the primary impact findings or to generate hypotheses about impacts. The primary outcomes and key subgroups are as follows:

- **Primary outcomes.** Primary evidence of the ERO programs' effect on high school outcomes is reflected by estimates of program impacts on students' performance in their core courses (GPA and credit accumulation). State test scores, while an important and policy-relevant outcome, are considered secondary indicators of program effectiveness due to inconsistencies in the availability of state test data across districts, and because the content of these tests differs across the states in the study. Student behaviors are also considered secondary outcomes, because while both ERO programs have components that address student behavior, they are not the focus of the programs. All subject-specific impact findings are also considered secondary. For example, while the estimated impact of the ERO programs on credit accumulation and GPA is presented for each core subject area separately (ELA, math, social studies, and science), this information is included only to contextualize the impact of the programs on all core subject areas combined.

- **Primary years.** Evidence of the ERO programs' effectiveness is determined by whether the programs have an impact on course performance in the program year or in the follow-up year. While school records data are also available for an additional follow-up year for Cohort 1, impacts in this year are considered secondary since they are based on a more restrictive sample of students.[80]

- **Primary subgroups.** The overarching question that guides the ERO study is whether the two reading programs together have an impact on student outcomes. Although impacts for other subgroups of students or schools (such as cohorts, reading interventions, or subgroups of students defined by ability or demographic characteristics) provide useful information about the relative

[79]These safeguards are based on the recommendations in Schochet (2008).
[80]Appendix I, which presents cohort-specific findings, provides impact findings for this year.

impact of supplemental literacy programs, they too are considered secondary indicators of effectiveness in this report.

This means that there are four primary indicators of program effectiveness in this report: estimated impacts on GPA and on credit accumulation, in both the program year and the follow-up year. As a second safeguard against "false positives," p-values for these four primary impact estimates were adjusted for multiple hypothesis testing.[81] Conclusions about the statistical significance of estimated impacts on these four primary high school outcomes are the same regardless of whether or not p-values are adjusted for multiple hypothesis testing. Therefore, for simplicity, the p-values in the impact tables in this report are unadjusted. Adjusted p-values for the four primary outcomes are located in Appendix F.[82]

Before presenting these new impact findings (Chapter 4), the next chapter provides an overview of the key findings from the two previous study reports. This includes a review of findings related to how the programs were implemented in the study schools, as well as the estimated impact of the programs on students' reading achievement and reading behaviors.

[81]P-values are adjusted based on the re-sampling methods described in Westfall and Young (1993). These methods take into account the fact that impact estimates (and hypothesis tests) for the program year and the follow-up year are not independent because they are based on overlapping samples of students. (Conversely, other methods assume that the hypothesis tests are independent and are therefore too conservative.) See Appendix F for details.

[82]Adjusted p-values were not calculated for impacts on the other outcomes in this report (state tests, attendance, etc.) — nor for impacts on student or school subgroups — because these analyses are secondary in this report, as noted earlier.

Chapter 3

Program Implementation and Impacts on Reading Achievement and Reading Behaviors

The first two reports from the Enhanced Reading Opportunities (ERO) study presented information on the implementation of the two ERO programs and their impacts on reading achievement and reading behaviors. Each report focused on one year of program implementation (2005-2006 or 2006-2007). The second report also presented some information comparing Year 1 and Year 2 implementation and student-level impacts. This chapter reviews the key findings from those reports and presents information from analyses of the same outcomes conducted on the sample of students pooled across both cohorts. The following key findings are discussed in more detail throughout this chapter:

- In the second year of implementation, 26 of the 34 participating high schools were implementing the ERO programs in a manner that was faithful to the program models.
- In the second year of the study, schools operated the ERO programs for almost the entire school year.
- In both implementation years, students in the ERO group received more literacy support than students in the non-ERO group.
- The ERO programs improved students' reading comprehension skills over the course of ninth grade.
- The ERO programs did not increase students' vocabulary test scores, nor did the programs affect how much students read or their use of the reading strategies taught in the two programs.

Program Implementation

The implementation research addressed the fidelity of program implementation as well as the extent to which the experience with literacy instruction was substantially different, in both degree and type, for the students in the ERO group compared with those in the non-ERO group. The evaluation also assesses whether the ERO group students received more and different literacy instruction than their counterparts outside the ERO program. First, the ERO study team measured how many hours of the programs the ERO students actually received.

Then, ERO and non-ERO group students' class schedules and self-reported participation in other literacy support services either in or outside school were analyzed.

Implementation Fidelity: Did the ERO Programs as Implemented Match the Intent of Their Design?

To measure the degree to which the observed operation of the ERO program in a given high school approximated the model specified by the program developers, data from classroom observations were used to calculate composite ratings of implementation fidelity. Study team members observed ERO classrooms during three separate site visits (one in the spring of the first year, and one in the fall and one in the spring of the second year), and composite measures of implementation fidelity were computed for each school. The protocols used in the classroom observations provided a structured process for observers to rate characteristics of the ERO classroom learning environments and the ERO teachers' instructional strategies. The instrument included ratings for six constructs (that is, program characteristics) that are common to both programs and ratings for seven program-specific constructs.

The analysis of the implementation fidelity sought to capture two key overarching dimensions of both programs: the classroom learning environment and the instructional strategies focused on reading comprehension. A composite measure of implementation fidelity for each dimension was calculated from the average ratings of both general and program-specific constructs. The ratings ranged from 1 to 3. An average rating of 2.0 or higher means the school's ERO program was "well aligned" with the developers' implementation specifications on a given dimension, while a rating of 1.0 to 1.9 means the program was "moderately or poorly aligned."[83]

The study team also looked at both implementation dimensions together for each school to get a sense of *overall* implementation fidelity. Because the classroom learning environments and reading 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 as well aligned. The overall implementation of the ERO program was rated as moderately or poorly aligned with the program model if at least one of the two key program dimensions was rated as moderately or poorly aligned.

[83]The ERO implementation fidelity analysis grouped schools into three categories: "well aligned," "moderately aligned," and "poorly aligned" with the program model. However, to prevent the risk of disclosure, the latter two categories are combined when presenting the results. For more detail on the classroom observation protocols, how the implementation fidelity measures were constructed, and scale reliability, see Appendix C.

Table 3.1 presents the distributions of schools (all schools and by program) by implementation fidelity categories for each implementation dimension and overall at the time of the spring site visit in the second year of implementation. These distributions represent where schools had "arrived" in terms of how well their ERO programs aligned with the developers' models by the end of the study's required implementation period. As can be seen in the table, 31 of 34 schools were rated as having well-aligned implementation in terms of learning environment, 28 of 34 schools were rated as well aligned in terms of reading comprehension instruction, and 26 of 34 schools had well-aligned implementation on both dimensions.

The rating 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 lower-fidelity programs effective.

The analyses of the classroom observation data from the two years of implementation resulted in the following key findings about implementation fidelity:[84]

- In general, the participating high schools implemented the ERO programs in a manner that was well aligned to the program models during the second year of implementation.
- Implementation was stronger in the second year than in the first year. The study team found that the ERO courses at 26 of the 34 high schools were categorized as well aligned with the program models in the second year, compared with 16 schools in the first year.[85]
- On average, implementation fidelity improved on both dimensions — learning environment and comprehension instruction — from the first year to the second year.

Figure 3.1 and Figure 3.2 show how the average ratings (on a scale from 1 to 3) for learning environment and comprehension instruction, respectively, got larger for both Reading Apprenticeship Academic Literacy (RAAL) and Xtreme Reading schools from the spring site

[84]Statistical tests of the difference in average fidelity ratings across implementation years and subgroups of schools were not conducted.

[85]The comparison of fidelity in Year 1/Year 2 focuses on the two *spring* site visits, so that the comparison is based on data gathered at the same time in both years of the study.

The Enhanced Reading Opportunities Study

Table 3.1

Number of ERO Classrooms with Well-, Moderately, or Poorly Aligned Implementation to Program Models on Each Implementation Dimension, by ERO Program: Year 2, Spring

Implementation Dimension	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Learning environment			
School average	2.46	2.63	2.28
Well-aligned implementation (composite rating is 2.0 or higher)	31	--	--
Moderately or poorly aligned implementation (composite rating is less than 2.0)	3	--	--
Comprehension instruction			
School average	2.33	2.27	2.38
Well-aligned implementation (composite rating is 2.0 or higher)	28	--	--
Moderately or poorly aligned implementation (composite rating is less than 2.0)	6	--	--
Combined dimensions			
School average	2.39	2.45	2.33
Well-aligned implementation on both dimensions	26	13	13
Moderately or poorly aligned implementation on at least one dimension	8	4	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 begninning stages of development in other areas. The implementation for these dimensions was designated as moderately aligned.

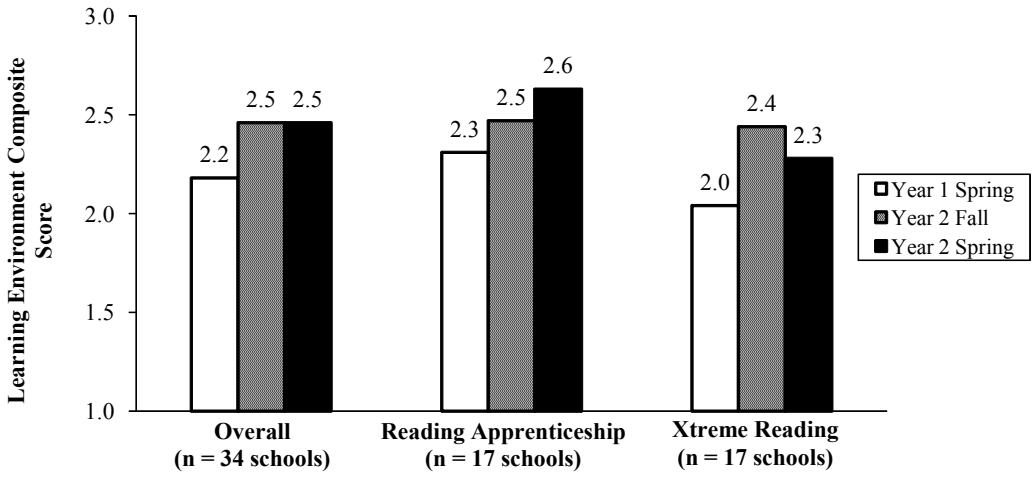
Implementation with scores of 2.0 or higher for a given dimension exhibited well-developed fidelity on several areas and at least moderate development in most other areas. The implementation for these dimensions was designated as well aligned.

Some counts have been suppressed to protect confidentiality.

The Enhanced Reading Opportunities Study

Figure 3.1

Learning Environment Composite Scores, by ERO Program



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.

Implementation with scores of 2.5 or higher for a given dimension exhibited well-developed fidelity on most areas and at least moderate development on most other areas. The implementation for these dimensions was designated as very well aligned.

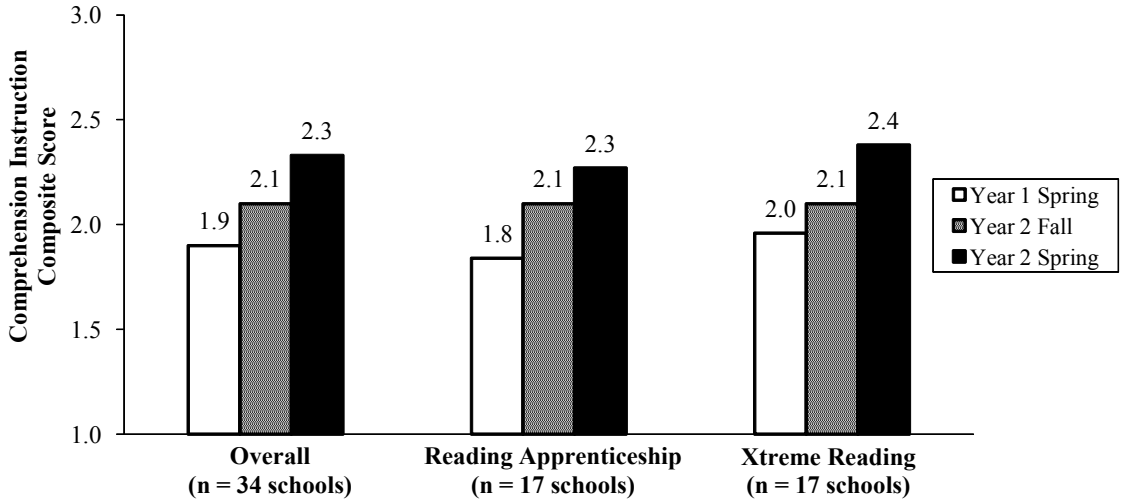
visit in Year 1 through the spring site visit in Year 2.[86] The study team found that, on average, the 25 teachers who returned to teach the ERO programs after having taught all of the first year

[86]In the spring of the first year of implementation, 26 schools were rated as well aligned on the learning environment dimension. By the spring of the second year, the number of schools considered well aligned to the program models increased to 31. At the spring site visit during the first year of implementation, 16 schools were considered as having well-aligned fidelity of implementation for the comprehension instruction dimension. The number of schools that had well-aligned implementation for comprehension instruction increased by the spring site visit of the second year, when 28 schools were rated as well aligned.

The Enhanced Reading Opportunities Study

Figure 3.2

Comprehension Instruction Composite Scores, by ERO Program



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.

Implementation with scores of 2.5 or higher for a given dimension exhibited well-developed fidelity on most areas and at least moderate development on most other areas. The implementation for these dimensions was designated as very well aligned.

had higher implementation ratings in the second year than in the first year.[87]

These teachers established learning environments that were better aligned to the program models (average rating of 2.5 in the second year and 2.4 in the first year). They had higher implementation fidelity ratings in terms of their comprehension instruction in their second year

[87]The section in Chapter 1 on program implementation and teacher training includes a discussion of returning and replacement teachers.

teaching the ERO programs than in their first year (average rating of 2.3 in the second year compared with 2.0 in the first year).

- The study team also found that, on average, in the second year of program implementation, the nine replacement teachers had higher implementation fidelity ratings in terms of both the learning environment established in their classrooms and their comprehension instruction (average ratings of 2.4 and 2.3, respectively) than the teachers they replaced (average ratings of 1.7 and 1.8, respectively).

Figure 3.3 plots the composite fidelity scores (averages of the learning environment and comprehension instruction ratings) across the two spring site visits from the two years of the study. The diagonal line indicates no change in implementation fidelity from Year 1 to Year 2. Implementation improved from Year 1 to Year 2 in schools above the diagonal line, and declined in schools below the diagonal line. The figure illustrates that these overall fidelity scores in the second year were the same as or higher than scores in the first year for 28 of the 34 schools (that is, on or above the diagonal line). Figure 3.3 also shows that, while there were nine replacement teachers at the beginning of the second year of the study, almost all of the schools where there was a replacement teacher had a higher rating in the second year than in the first year (that is, above the diagonal line). Figure 3.3 also shows that at 19 of the 25 schools where the ERO teachers returned after having taught all of the first year, the average implementation composite ratings were the same or higher in the second year.[88]

Thus, across the two years of implementation in the study, the ERO programs as implemented moved closer to the intent of the programs' designs in terms of classroom learning environments and comprehension instruction practices. Implementation fidelity ratings were higher in the second year compared with ratings in the first year, both in schools where replacement teachers taught the programs in the second year and in schools where the programs were taught for a second full year by the same teacher. Also, more schools in the second year achieved ratings of "well aligned" on both key program dimensions.[89]

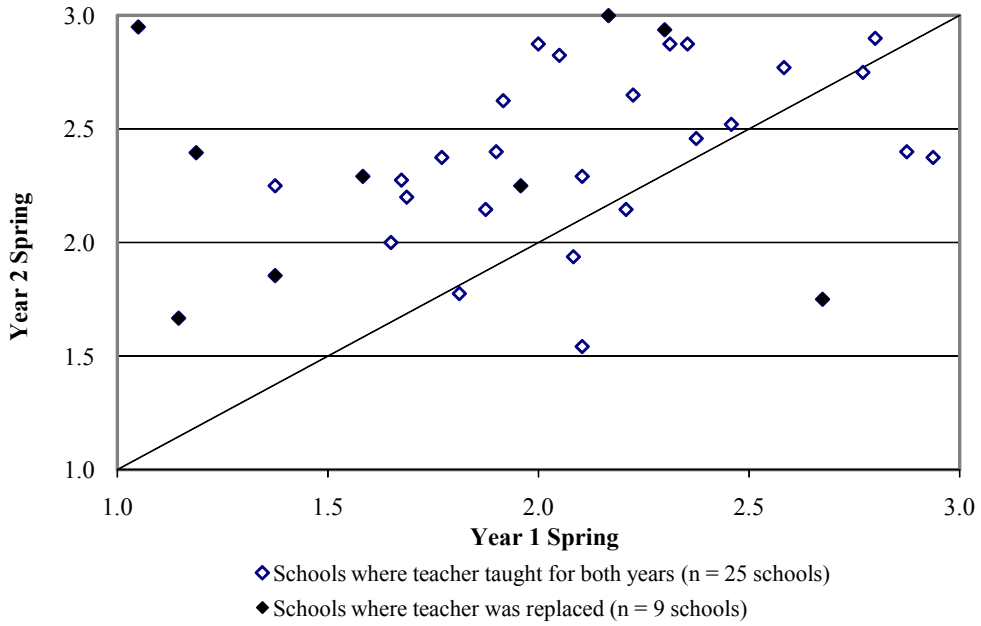
[88]Tables C.4 and C.5 in Appendix C show implementation fidelity scores by teachers' experience with the ERO program (that is, for the 9 schools where the ERO teacher was replaced compared with the 25 schools where the ERO teacher taught in both years).

[89]The implementation fidelity findings for implementing each of the two programs, RAAL and Xtreme Reading, are similar across both years of implementation. Details about implementation fidelity findings by program can be found in Appendix C.

The Enhanced Reading Opportunities Study

Figure 3.3

Composite Fidelity Scores, by Site Visit



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.

Implementation with scores of 2.5 or higher for a given dimension exhibited well-developed fidelity on most areas and at least moderate development on most other areas. The implementation for these dimensions was designated as very well aligned.

The diagonal line represents points where the Year 1 fidelity score is equal to the Year 2 fidelity score. For schools located above the diagonal line, the Year 2 fidelity score is greater than the Year 1fidelity score. For schools located below the diagonal line, the Year 2 fidelity score is lower than the Year 1 fidelity score.

Program Duration and Student Attendance: How Much ERO Instruction Did ERO Students Receive?

Based on the ERO programs' theory of action, the impact is hypothesized to be a function, in part, of how much exposure the ERO students have to the ERO classes throughout the school year. The evaluation team collected data about how long the ERO classes ran during the school year, how frequently the classes met, and whether and how often students attended these courses to clarify whether or not ERO students were actually exposed to the entire program. These data provide an indication of the overall "dosage" of the ERO interventions that students in the ERO group received.

- In their first year of operation, ERO classes started about six weeks into the school year on average and met for an average of 7.7 months. In the second year, the ERO courses were in session for an average of one and a half months longer than the first year, beginning just over two weeks after the start of the school year on average and operating for an average of 9.1 months of the academic year. (See Table 3.2.)

The difference in duration is explained by differences in the timing of student recruitment and random assignment each year. During the first year, most schools recruited and randomly assigned students during the first months of the school year, delaying the start of ERO courses. During the second year, many schools were able to recruit students before the school year began, allowing many ERO programs to start on time. Given that the curricula for the ERO classes were designed to be covered in a full school year,[90] the shorter duration in the first year suggests that many teachers did not have the time necessary to teach the full ERO course.

- On average, students randomly assigned to take the ERO course at their high school were able to take advantage of the opportunity, attending throughout the school year. Over 90 percent of the students randomly assigned to take the ERO course at their high school attended the course at least once, and the majority of these students were still attending at the end of the year.

Among the students in the first cohort, 95.5 percent ever attended an ERO class, while 91.3 percent of students in the second cohort ever attended class. Of the first cohort ERO students, 91.2 percent were attending the course the last month of school compared with 86.9

[90] Across schools, a school year lasted approximately 9 to 10 calendar months, depending on district and school calendars.

The Enhanced Reading Opportunities Study

Table 3.2

Attendance in ERO Classes,
GRADE Respondent Sample in the ERO Group

Characteristic	All Schools		Reading Apprenticeship Schools		Xtreme Reading Schools	
	Cohort 1	Cohort 2	Cohort 1	Cohort 2	Cohort 1	Cohort 2
Ever attended an ERO class during the year (%)	95.5	91.3	94.9	92.1	96.0	90.5
Attending ERO classes at the end of the year (%)	91.2	86.9	91.0	88.2	91.4	85.6
Average daily attendance rate in ERO classes per month (%)	82.7	79.3	81.7	80.0	83.6	78.6
Average number of months ERO program was in operation	7.7	9.1	7.8	9.1	7.7	9.0
Average number of months attending ERO classes	7.1	8.0	7.1	8.1	7.1	7.8
Average number of hours ERO class met per month	13.6	13.6	13.5	13.8	13.7	13.5
Average number of hours student attended ERO class per month	11.3	10.8	11.2	11.1	11.5	10.6
Average number of hours student attended ERO during the school year[a]	87.0	98.2	87.4	100.8	88.6	95.5
Sample size	1,408	1,264	686	645	722	619

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

NOTE: [a]The average number of hours student attended ERO during the school year is the average number of months ERO program was in operation multiplied by the average number of hours student attended class per month.

percent of the second cohort of ERO students. Since most of the recruitment of the second cohort of students happened before the start of the school year (at the end of the students' eighth-grade year) and more students are likely to leave the district in the summer between middle and high school than while classes are in session, more students in the second cohort left the district between random assignment and the start of the program.

- ERO students attended class regularly. The average daily attendance rate for the ERO classes was 82.7 percent for the first cohort and 79.3 percent for the second cohort.

On average, ERO students in the first year received 11.3 monthly hours of ERO instruction out of 13.6 hours offered. ERO students in the second cohort received 10.8 monthly hours of ERO instruction out of 13.6 hours offered. Given that the programs were designed to offer 225 minutes of weekly instruction (15 hours per month), the cohorts received 75.3 percent and 72.0 percent of intended instruction, respectively.

- Students in the second cohort received more ERO instruction overall given the longer duration of the ERO classes in the second year.

Table 3.2 presents multiple measures of the dosage of the intervention — the amount of literacy support — that the ERO group students received through the ERO classes. Even though they had lower attendance rates, students in the second cohort received an average of 98.2 total annual hours of ERO instruction compared with 87.0 hours for the students in the first cohort.

Student Course Schedules and Literacy Support Services: How Different Were the ERO Courses from "Business as Usual"?

The ERO study team also hypothesized that the impact of the ERO program is a function of the contrast between the ERO dosage (that is, the literacy support provided by the ERO classes) and the amount of literacy support available to non-ERO students. Impacts of the ERO program could be falsely minimized if the ERO course replaced another literacy-focused course taken by the majority of non-ERO students. On the other hand, impacts could be falsely enhanced if ERO group students were taking another literacy-focused course beyond ERO that non-ERO students were not taking. Even if both ERO and non-ERO students are taking similar courses outside the ERO programs, if all students are taking literacy-rich English language arts (ELA) courses or supplemental literacy courses, the impacts for the program could be minimized because the ERO class would not be adding much value to an already literacy-rich environment.

Several types of data were collected and analyzed to measure the contrast in the literacy support available and received by ERO and non-ERO group students. The data on ERO class attendance provided 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.[91] Student schedule data were used to assess course-taking patterns of students participating in the study, providing an understanding of how the ERO courses fit into student schedules and an opportunity to explore potential contrasts in the enrollment of ERO students and non-ERO students in different types of core content (ELA, mathematics, science, and social studies) and noncore content courses, including electives. Students were also surveyed about the frequency and duration of their participation in classes or tutoring services in school and outside of school that aimed to improve reading and writing skills. These data are available for students in both the ERO and the non-ERO groups and are

[91]The ERO classes were not a source of literacy support for non-ERO students: 9 out of the 1,241 students in the non-ERO group in the first year and 17 out of the 1,150 students in the non-ERO group in the second year were enrolled in the ERO classes.

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.

Key findings from the first two reports from the ERO study about the availability of literacy support services for ERO and non-ERO students are presented below:

- The main difference in students' schedules for ERO and non-ERO group students was that the ERO course replaced a nonliteracy-focused elective course for most students.

In both the first and second years, ERO students were typically enrolled in the same core courses as non-ERO students and a similar variety of elective courses, with the exception of the ERO group students substituting the ERO program for one of these electives. With few exceptions, all ERO and non-ERO students were enrolled in one ELA course and one mathematics course (on average, both ERO and non-ERO students took 1.01 courses in both these subjects), and almost all students, regardless of participation in ERO, took a history and a science course in ninth grade (on average, ERO and non-ERO students took between 0.94 and 0.97 course in these subjects). Other than the ERO course, students took a similar variety of noncore courses in areas such as career and technical services, visual and performing arts, physical education and health, foreign languages, and support services.[92]

- Students who were randomly assigned to the study's ERO group reported a higher frequency of participation in supplemental literacy services than students in the non-ERO group in both years of implementation. (See Figure 3.4.)

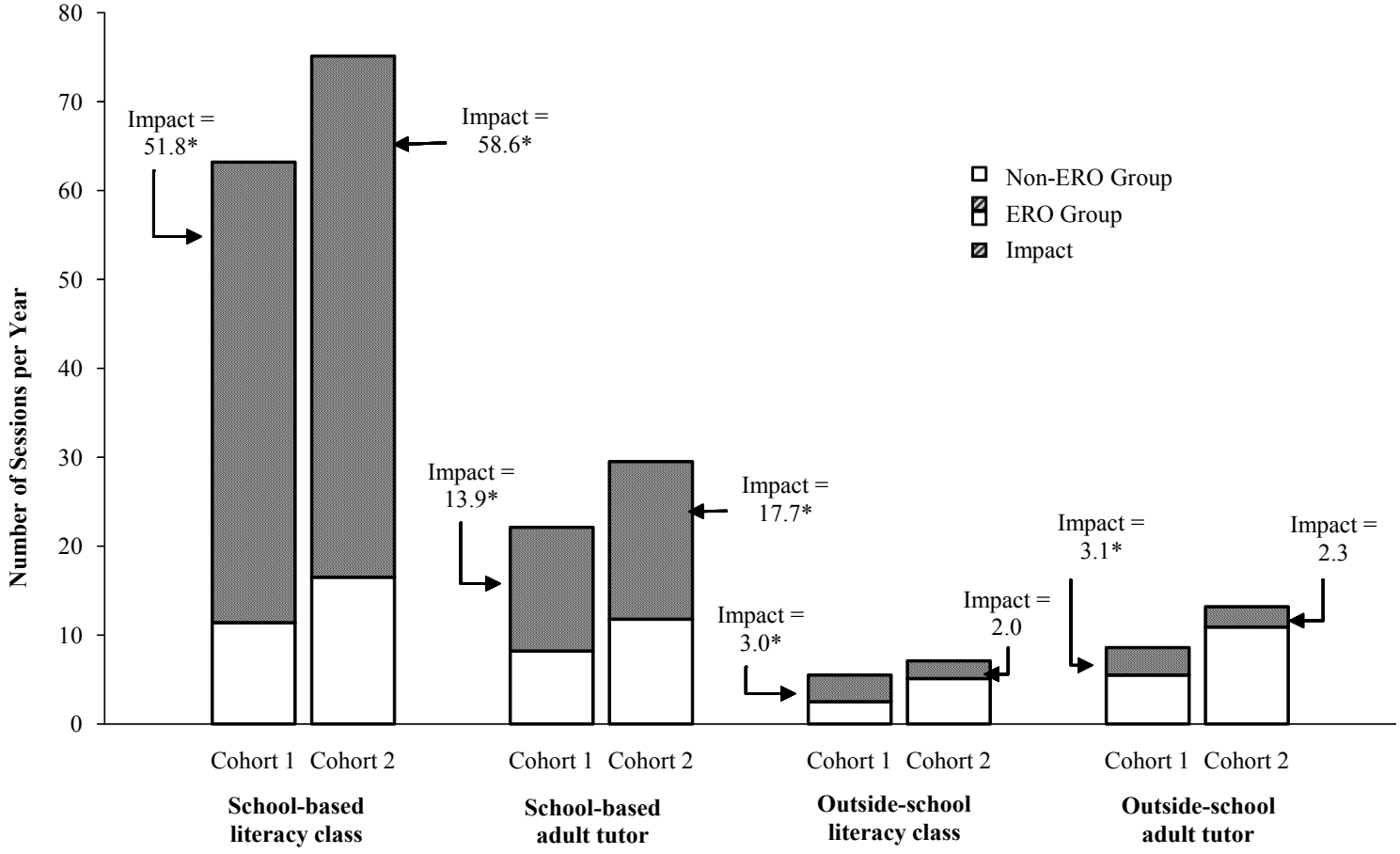
Both cohorts of ERO students reported participating in more sessions of school-based literacy classes than non-ERO students in that cohort (51.8 more sessions for the first cohort and 58.6 more sessions for the second cohort).[93] These differences are statistically significant. Although the students were not specifically instructed to include the ERO class in their survey

[92]See Corrin et al. (2008), Table 4.3, for detailed information on students' course taking during the second year of ERO implementation. The analysis of course schedules was not conducted with the same amount of detail during the first year of implementation.

[93]As described in Chapter 2, the measures of student participation in literacy support activities were constructed from items on a student survey administered in the spring of ninth grade to ERO and non-ERO students. Appendix B contains information about how these measures were constructed, as well as a copy of the survey.

The Enhanced Reading Opportunities Study

Figure 3.4
Participation in Supplemental Literacy Support Activities



Cohort 1 sample size = 2,412 students
Cohort 2 sample size = 2,161 students

Figure 3.4 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities student survey administered at the end of students' ninth-grade year (spring 2006 for Cohort 1 and spring 2007 for Cohort 2).

NOTES: 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 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 5.2 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

response, it is likely that their responses reflect the ERO classes they were taking. For this reason, the study team interprets this impact as confirmatory.[94]

Thus, when looking across all three categories of implementation data — implementation fidelity, program "dosage," and the contrast in literacy support received by ERO students compared with non-ERO students — it appears that, by the second year of implementation, many schools were able to implement full-year programs faithful to the design of the developers that offered literacy support different from what the non-ERO students were receiving. These findings suggest that the programs were given a strong or "fair" test of effectiveness, particularly in the second year of implementation. But did these programs make a difference in terms of students' reading outcomes; that is, were they effective?

Program Impacts on Reading Achievement and Reading Behaviors

Because the study's two supplemental literacy programs —RAAL and Xtreme Reading — focus on producing immediate improvements in students' reading comprehension ability, the impact analyses presented in the first two study reports placed a higher priority on reading comprehension test scores. 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 strategiesthat are characteristic of proficient readers. For this reason, the analyses also examined impacts on vocabulary test scores and on three measures of students' reading behaviors.

Impacts on Student Reading Achievement: Did the ERO Programs Improve Students' Reading Skills?

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. This section includes a description of the key impact findings for both cohorts of participating students pooled together.[95] Although implementation of the program was stronger during the second year than

[94]ERO students in both cohorts also reported participating in more sessions with a school-based literacy tutor than their non-ERO counterparts (13.9 more sessions for the first cohort and 17.7 more sessions for the second cohort). These differences are also statistically significant. Both cohorts of ERO students reported participating in more literacy support activities outside of school than non-ERO students. These findings were statistically significant for the first cohort of students (3.0 more sessions in a literacy class and 3.1 more sessions with a literacy tutor), but were not statistically significant for the second cohort of students.

[95]See Chapter 2 for a discussion of the sample of students used to estimate impacts on reading achievement ("GRADE respondent sample").

the first, impacts on reading achievement do not differ by a statistically significant amount across the two cohorts. This indicates that it is appropriate to combine the impact findings for the two cohorts, as is done in this section. However, separate findings for each cohort — which were presented in the prior two reports — are also described, along with the tests of the differences in impacts between the two cohorts.

Table 3.3 presents the impacts of the ERO programs on students' GRADE test scores for both cohorts combined.

- Averaged across all 34 participating high schools and both cohorts of ninth-grade students, the ERO programs improved reading comprehension test scores by 0.9 standard score point. This impact is statistically significant (effect size = 0.09; p-value = 0.002).[96]

Table 3.3 shows that this impact corresponds to an improvement from the twenty-third percentile nationally, as represented by the average scores for students in the non-ERO group, to the twenty-fifth percentile nationally, as represented by the average scores for students in the ERO group. Both the ERO and non-ERO students were still reading far below grade level at the end of ninth grade. The ERO students' average reading comprehension score of 90.1 corresponds to a grade equivalent of 6.1 (that is, the first month of sixth grade). The non-ERO students' average score of 89.2 corresponds to a grade equivalent of 5.9 (that is, the last month of fifth grade). This suggests that both the ERO and non-ERO students, who were tested close to the end of ninth grade, were about four years below grade level at the end of the program.

- The estimated impact of the ERO programs on reading comprehension test scores across the 34 high schools was also statistically significant for each cohort separately: 0.9 standard score point (effect size = 0.09; p-value = 0.019) in the first year and 0.8 standard score point (effect size = 0.08; p-value = 0.042) in the second year.[97] The impact on students in the first cohort is not statistically significantly different from the impact on students in the second cohort.

Table 3.4 presents impact findings separately for each of the two reading interventions (RAAL and Xtreme Reading):

[96]As discussed in Chapter 2, the estimated impact effect size is expressed as a proportion of the overall variability (standard deviation) of the non-ERO group for a given outcome measure in the relevant year (program year or follow-up year). Confidence intervals for the impacts on GRADE reading achievement scores for the pooled sample and for each program separately can be found in Table F.1, in Appendix F.

[97]See Table 5.1 in Kemple et al. (2008) and Table 5.1 in Corrin et al. (2008) for detailed findings.

The Enhanced Reading Opportunities Study

Table 3.3

Impacts on Reading Achievement, GRADE Respondent Sample, Two Cohorts Combined

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools					
Reading comprehension					
Average standard score	90.1	89.2	0.9*	0.09*	0.002
Corresponding grade equivalent	*6.1*	*5.9*			
Corresponding percentile	*25*	*23*			
Reading vocabulary					
Average standard score	93.5	93.3	0.1	0.01	0.640
Corresponding grade equivalent	*7.8*	*7.7*			
Corresponding percentile	*32*	*32*			
Sample size	2,672	1,912			

SOURCE: MDRC calculations from the Enhanced Reading Opportunities GRADE assessment, administered at the end of ninth grade (spring 2006 for Cohort 1 and spring 2007 for Cohort 2).

NOTES: The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline and whether a student was overage for grade at random assignment. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation for the non-ERO group in the program year (all schools).

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 national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

- The estimated impact of the ERO programs on reading comprehension test scores in the 17 high schools implementing the RAAL program was 1.2 standard score points for both cohorts combined (effect size = 0.12; p-value = 0.002).

The Enhanced Reading Opportunities Study

Table 3.4

Impacts on Reading Achievement,
GRADE Respondent Sample, Two Cohorts Combined, by Program

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
Reading Apprenticeship schools					
Reading comprehension					
Average standard score	90.0	88.8	1.2*	0.12*	0.002
Corresponding grade equivalent	*6.1*	*5.8*			
Corresponding percentile	*25*	*22*			
Reading vocabulary					
Average standard score	93.3	93.3	0.0	0.00	0.995
Corresponding grade equivalent	*7.7*	*7.7*			
Corresponding percentile	*32*	*32*			
Sample size	1,331	924			
Xtreme Reading schools					
Reading comprehension					
Average standard score	90.2	89.7	0.6	0.05	0.165
Corresponding grade equivalent	*6.1*	*6.0*			
Corresponding percentile	*25*	*24*			
Reading vocabulary					
Average standard score	93.6	93.3	0.3	0.03	0.517
Corresponding grade equivalent	*7.8*	*7.7*			
Corresponding percentile	*32*	*32*			
Sample size	1,341	988			

SOURCE: MDRC calculations from the Enhanced Reading Opportunities GRADE assessment, administered at the end of ninth grade (spring 2006 for Cohort 1 and spring 2007 for Cohort 2).

NOTES: The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline and whether a student was overage for grade at random assignment. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation for the non-ERO group in the program year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

To contextualize the magnitude of this impact, Table 3.4 shows that, on average, students in the non-ERO group scored at the twenty-second percentile nationally, while RAAL students in the ERO group scored at the twenty-fifth percentile nationally.

For the first cohort of students, the estimated impact of the ERO programs on reading comprehension test scores in the 17 high schools implementing the RAAL program was 0.9 standard score point and not statistically significant (effect size = 0.09; p-value = 0.097). For the second cohort, the estimated impact was 1.4 standard score points and statistically significant (effect size = 0.14; p-value = 0.015).[98] These impacts are not statistically significantly different from each other.

- The estimated impact of the Xtreme Reading program on reading comprehension test scores was 0.6 standard score point for both cohorts combined, but this impact is not statistically significant (p-value = 0.165). This impact represents an effect size of 0.05.

Table 3.4 shows that, on average, students in the non-ERO group scored at the twenty-fourth percentile nationally, while Xtreme Reading students in the ERO group scored at the twenty-fifth percentile nationally.

For the first cohort of students, the estimated impact of the ERO programs on reading comprehension test scores in the 17 high schools implementing the Xtreme Reading program was 0.9 standard score point (effect size = 0.09; p-value = 0.090). For the second cohort, the impact was 0.2 standard score point (effect size = 0.02; p-value = 0.672).[99] Neither impact estimate is statistically significant, nor are they statistically significantly different from each other.

- While the two programs have different patterns of impacts on reading comprehension test scores, the difference between these impacts is not statistical ly significant. Therefore, it cannot be concluded that RAAL produced a different impact from the one that Xtreme Reading produced.

This finding suggests that it is appropriate to estimate the impact of the two programs together as a class of intervention as done in Table 3.3 and for the impact findings on high school impacts found in Chapter 4. But, in both cases, impacts by program are also included for informational purposes.

- Considered together or separately, the ERO programs did not have statistically significant impacts on vocabulary test scores.

[98]See Table 5.2 in Kemple et al. (2008) and Table 5.2 in Corrin et al. (2008) for detailed findings.
[99]See Table 5.2 in Kemple et al. (2008) and Table 5.2 in Corrin et al. (2008) for detailed findings.

Table 3.3 shows that there was not a statistically significant impact of the programs overall on vocabulary test scores (p-value = 0.640) for the two cohorts of ninth-grade students combined. The vocabulary test scores averaged across all 34 participating high schools were similar for the ERO and non-ERO groups — a scale score of 93.5 for the ERO students and 93.3 for the non-ERO students. Table 3.4 shows that neither ERO program produced a statistically significant impact on vocabulary test scores for the two cohorts of students combined. In the RAAL schools, the ERO group scored the same as the non-ERO group; in the Xtreme Reading schools, the ERO group scored 0.3 standard score point higher than the non-ERO group. There are no statistically significant findings on reading vocabulary for either of the two cohorts analyzed separately.[100]

Impacts on Student Reading Behaviors: Did the ERO Programs Change How and How Much Students Read?

Each of the two supplemental literacy programs seeks to motivate students to read more. They do this both by providing opportunities for students to read and to 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. The impact findings presented in this section of the chapter focus on three measures of reading behavior that were derived from the student survey administered at the end of ninth grade: the amount of school-related reading, the amount of non-school-related reading, and the use of reflective reading strategies (that is, the use of reading strategies in which students reflect on what they are reading and ask questions about the text to better understand what they read). As explained in Chapter 2, the measures of reading behavior were developed from the survey that was administered to students in the study sample at the end of their ninth-grade year.[101]

- Overall, the two ERO programs, together or separately, were not effective in changing students' reading behaviors in either year.[102]

[100]See Tables 5.1 and 5.2 in Kemple et al. (2008) and Tables 5.1 and 5.2 in Corrin et al. (2008) for detailed findings on each cohort separately.

[101]A list of the survey items used to create these three measures and the survey itself are presented in Appendix B.

[102]The ERO programs did produce a positive and statistically significant impact on student-reported use of reflective reading strategies in the second year of implementation (effect size = 0.09; p-value = 0.033). However, a composite qualifying statistical test for multiple hypothesis tests across all three reading behaviors in Year 2 was not statistically significant, suggesting that this single impact finding should be interpreted with caution. See Chapter 5 in the second ERO study report (Corrin et al., 2008).

For this report, analyses of the impacts on the three reading behavior outcomes pooled across both cohorts were conducted. Table 3.5 and Table 3.6 present the impact findings for these outcomes overall and by program. None of the impacts is statistically significant.

This chapter has reviewed what was learned about the operation of the supplementary literacy interventions during their two years of implementation, as well as their impacts on students' reading test scores and reading behaviors. The next chapter of the report presents what else has been learned about the effectiveness of the ERO programs by analyzing their impact on school-based student achievement and behavior outcomes. These outcomes — course grades, credit earning, state assessment scores, disciplinary actions, and attendance — represent measures more typically used by schools and districts to assess how their students are doing, and may be considered by many readers to be more policy-relevant than the more proximal reading outcomes analyzed for the first two reports.

The Enhanced Reading Opportunities Study

Table 3.5

Impacts on Reading Behaviors,
GRADE Respondent Sample, Two Cohorts Combined

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools					
Amount of school-related reading (prior-month occurrences)	45.2	44.6	0.6	0.01	0.617
Amount of non-school-related reading (prior-month occurrences)	28.5	27.6	0.9	0.03	0.330
Use of reading strategies (4-point scale)	2.7	2.7	0.02	0.04	0.199
Sample size	2,670	1,903			

SOURCE: MDRC calculations from the Enhanced Reading Opportunities student survey, administered at the end of the program implementation year (spring 2006 for Cohort 1 and spring 2007 for Cohort 2).

NOTES: The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline and whether a student was overage for grade at random assignment. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation for the non-ERO group in the program year (all schools).

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 2.5 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

The Enhanced Reading Opportunities Study

Table 3.6

Impacts on Reading Behaviors, GRADE Respondent Sample, Two Cohorts Combined, by Program

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
Reading Apprenticeship schools					
Amount of school-related reading (prior-month occurrences)	47.1	48.4	-1.3	-0.03	0.456
Amount of non-school-related reading (prior-month occurrences)	29.0	29.5	-0.6	-0.02	0.679
Use of reflective reading strategies (4-point scale)	2.7	2.7	0.0	0.01	0.799
Sample size	1,331	921			
Xtreme Reading schools					
Amount of school-related reading (prior-month occurrences)	43.3	40.8	2.5	0.06	0.131
Amount of non-school-related reading (prior-month occurrences)	28.0	25.7	2.4	0.07	0.071
Use of reflective reading strategies (4-point scale)	2.7	2.6	0.0	0.06	0.122
Sample size	1,339	982			

SOURCE: MDRC calculations from the Enhanced Reading Opportunities student survey, administered at the end of the program implementation year (spring 2006 for Cohort 1 and spring 2007 for Cohort 2).

NOTES: The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline and whether a student was overage for grade at random assignment. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation for the non-ERO group in the program year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

For each of the above measures, data are missing for no more than 2.7 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

Chapter 4

Impacts on High School Outcomes

The primary focus of the Enhanced Reading Opportunities (ERO) evaluation is to assess the impact of two supplemental literacy interventions — Reading Apprenticeship Academic Literacy (RAAL) and Xtreme Reading — on adolescent students' reading comprehension skills and their overall academic performance during high school. As discussed in Chapter 1 and Chapter 3, the first two reports measured the impact of the interventions on students' reading comprehension skills and behaviors. This third and final report focuses on the programs' impacts on students' school-based outcomes. With that focus, the impact findings presented in this chapter address two questions:

- What are the impacts of the supplemental literacy programs on students' performance in their courses and achievement on state assessments while they are participating in the programs and during the subsequent school year?
- What are the impacts on students' behaviors in school while they are participating in the programs and during the subsequent school year?

The findings in this chapter show that, overall, the ERO programs had a positive impact on multiple high school outcomes during the year that students were participating in the program (students' ninth-grade year). However, the programs did not generally produce statistically significant impacts on these outcomes during the following school year (most students' tenth-grade year).[103] More specifically:

- The ERO programs had a positive impact on students' course performance in core subject areas during the program year. Impacts on course performance for each of the two reading programs separately tell a similar story to the overall findings. Both RAAL and Xtreme Reading had a positive impact on students' grade point average (GPA) in core subject areas during the program year.
- Together, the two ERO programs also improved students' performance on English language arts (ELA) and mathematics state tests during the program year. Together, the ERO programs did not affect students' school behaviors during the program year.

[103] Adjusting the p-values for multiple hypothesis testing does not lead to different conclusions about whether key impacts on high school outcomes are statistically significant.

- In the school year following students' participation in the ERO programs, the programs, pooled together, did not have an impact on students' academic performance or school behaviors.
- Findings from subgroup analyses do not offer conclusive evidence that the programs were more effective for one subgroup of students than another.

This chapter presents impact findings for students in the study sample who have school records data (see Chapter 2 for a description of this sample).[104] This chapter begins by presenting impacts on school-based outcomes for both ERO programs pooled together. As described in Chapter 1, RAAL and Xtreme Reading have common goals for adolescent literacy development and share many instructional principles. For this reason, the focus of this report is on the findings for both programs together as a class of intervention.[105] Impacts on students' GRADE scores as well as their GPA and credit accumulation do not differ by a statistically significant amount across the two programs, suggesting that it is appropriate to estimate the impact of the two programs together. The chapter then presents findings for each of the two ERO programs separately for informational purposes.

Overall Impacts

This section presents impacts on school-based outcomes across all 34 high schools in the evaluation. These results, which are pooled across the two reading programs, provide evidence about the effectiveness of supplemental literacy interventions of the type represented by RAAL and Xtreme Reading. This section begins by discussing the programs' impact on students' performance in their core classes. This is followed by a presentation of impacts on students' performance on state tests and on their school behaviors.

---

[104]As discussed in Chapter 3, the estimated impact of the ERO programs on students' reading comprehension scores does not differ by a statistically significant amount across the two cohorts (effect size = 0.09 for Cohort 1 and 0.08 for Cohort 2). Hence, the decision was made to present impacts on high school outcomes across both cohorts pooled together. Appendix I presents impacts on the key academic performance outcomes for each cohort separately, including impacts for an additional follow-up year for Cohort 1. As discussed in that appendix, differences in impacts between the two cohorts on key outcomes are not statistically significant, suggesting that it is appropriate to pool the two cohorts.

[105]By pooling the sample across all schools in the study, the analysis can also detect impacts that are smaller than those that can be detected for each ERO program separately.

Course Performance

One of the goals of the reading interventions is to improve students' performance in their core courses. In this study, students' performance in their core courses is measured using their GPA and credit accumulation in core classes (ELA, social studies, science and mathematics) during the program year and during the school year following program participation.[106]

GPA

Table 4.1 presents the estimated impacts of the ERO programs on students' GPA in their core classes.[107] The first panel of this table presents impacts on students' GPA in ninth grade, the year of the program, while the second panel presents impact findings for the follow-up year which is most students' tenth-grade year. To help better understand the composition of the programs' impact on all core subject areas combined, Table 4.1 also displays impacts on students' grades in each core subject area separately.

- Together, the ERO programs had a positive impact of 0.06 point on a four-point scale on students' GPA in core subject areas during the program year (effect size = 0.07; p-value = 0.002).[108] However, the programs did not have a statistically significant impact on core GPA during the follow-up year.[109]

As shown in the first panel of Table 4.1, ERO students had an average GPA of 1.60 points while non-ERO students had an average GPA of 1.53 points during the program year. Thus, the ERO programs had an impact of 0.06 point on students' GPA in core subject areas (effect size = 0.07; p-value = 0.002).[110] Adjusting the p-value for multiple hypothesis testing does not change the statistical significance of this finding.

Figure 4.1 helps to put the magnitude of the programs' impact on core GPA into perspective. As shown in the top portion of this figure, on average, neither the ERO nor non-ERO

---

[106]See Appendix F for figures comparing the distribution of GPA and credit accumulation in core courses for the ERO group and the non-ERO group.

[107]Confidence intervals for the impacts on GPA in the program and follow-up year can be found in Table F.2, in Appendix F.

[108]As discussed in Chapter 2, the estimated impact effect size is expressed as a proportion of the overall variability (standard deviation) of the non-ERO group for a given outcome measure in the relevant year (program year or follow-up year).

[109]This pattern of results is not simply due to the fact that the sample size is smaller in the follow-up year. A sensitivity analysis shows that estimated impacts in the program year for the subset of students who have school records in both the program year *and* the follow-up year (longitudinal sample) are positive and statistically significant, just as they are for the full school records sample. See Appendix H for these findings.

[110]Rounding causes the discrepancy in the difference.

The Enhanced Reading Opportunities Study

Table 4.1

Impacts on Grade Point Average (GPA), School Records Sample

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools					
Program year					
GPA in core subject areas	1.60	1.53	0.06*	0.07*	0.002
English language arts (ELA)	1.71	1.64	0.07*	0.06*	0.009
Social studies	1.66	1.58	0.09*	0.07*	0.003
Science	1.54	1.46	0.08*	0.07*	0.004
Math	1.50	1.46	0.04	0.03	0.177
Sample size	2,937	2,213			
Follow-up year					
GPA in core subject areas	1.59	1.54	0.04	0.05	0.061
English language arts (ELA)	1.77	1.68	0.09*	0.08*	0.005
Social studies	1.63	1.60	0.03	0.02	0.426
Science	1.52	1.47	0.06	0.05	0.077
Math	1.44	1.43	0.01	0.01	0.769
Sample size	2,542	1,894			

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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 sample sizes reported in this table are for students with school records data in the relevant year. Because students may not have earned grades in all core subject areas in a given year, sample sizes differ for impacts in the specific core subject areas.

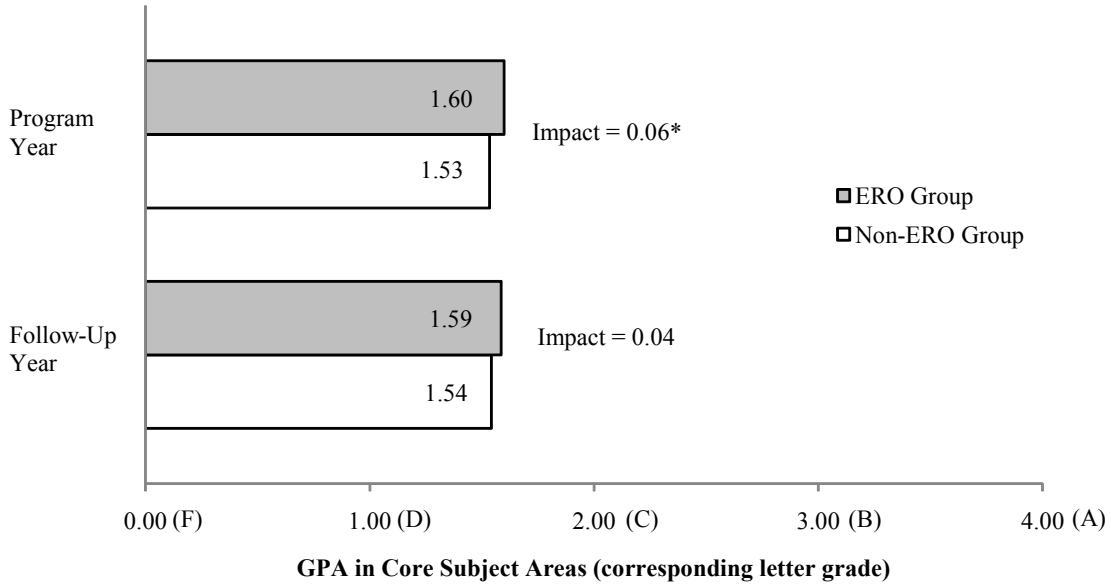
Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The Enhanced Reading Opportunities Study

Figure 4.1

Impacts on Grade Point Average (GPA), School Records Sample



SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group values are the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment.

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.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

group attained a C average in their core courses (that is, a GPA of 2.0 points). Yet, a C average in core courses is an important benchmark, since this grade is used by a majority of the study schools to determine "average" or "satisfactory" achievement in a given course.[111] Some of the districts also use this benchmark in determining whether students will be allowed to participate in extracurricular activities.[112] An average grade of C is also a predictor of the likelihood that students will graduate.[113] As displayed in the top portion of Figure 4.1, had students in the ERO group not been offered the opportunity to enroll in the supplemental reading class, they would have had an average GPA of 1.53 points in core classes (as represented by the GPA of students in the non-ERO group), which is 0.47 point below a C average.[114] This means that the reading programs' impact on GPA in core subjects (0.06 point) is such that these students' grades improved by 13 percent of what they needed to attain a C average.[115]

As seen in Table 4.1, this impact appears to be driven by the ERO programs' effect on students' grades in ELA, social studies and science. The programs had a positive and statistically significant impact of 0.07 point on students' grades in ELA (effect size = 0.06; p-value = 0.009), 0.09 point on students' grades in social studies (effect size = 0.07; p-value = 0.003) and 0.08 point on students' grades in science (effect size = 0.07; p-value = 0.004) during the program year. The estimated impact of the programs on students' grades in math is not statistically significant. This last finding is consistent with expectations that math abilities would be the least likely to be affected since it is the least reading-intensive core subject area and receives the least attention in the RAAL and Xtreme Reading programs.

---

[111]Five of the districts representing 18 schools refer to a C grade in a course as representing "average" or "satisfactory" performance explicitly in their course catalogs or student handbooks.

[112]Thirteen schools in the study require students to have a 2.0 GPA in order to participate in extracurricular activities (the other 21 schools have eligibility requirements related to the number of failing course grades). The National Collegiate Athletic Association (NCAA) also requires that incoming college freshman who plan to compete in athletics have a GPA of 2.0 or higher in core high school courses (NCAA, 2008).

[113]According to Allensworth and Easton (2007), almost three-fourths of Chicago Public Schools students with a C average at the end of their first year of high school graduated in four years. Specifically, they report the following four-year graduation rates by GPA: 93 percent for students with a 3.0 (rounded to the nearest 0.5), 86 percent for students with 2.5, 72 percent for students with a 2.0, 53 percent for students with a 1.5, and 28 percent for students with a 1.0. (These percentages are based on the unweighted averages for all credit-bearing classes.). As seen here, graduation rates are 36 percentage points higher for students with a C average (2.0) than for students with a D average (1.0).

[114]This is calculated as 2.0 points (a C average) minus 1.53 points.

[115]The value of 13 percent was calculated by dividing the impact (0.06 point) by the improvement in GPA needed to attain a C average (0.47 point).

The ERO programs did not affect students' GPA in core classes in the school year following their participation in the supplemental classes.[116] The second panel of Table 4.1 shows that average GPA in core courses for ERO students during the follow-up year was 1.59 points, while the average GPA for Non-ERO students during that year was 1.54 points.[117] The estimated impact of 0.04 point is not statistically significant, and therefore, it cannot be concluded that the programs had an impact on course performance in the follow-up year.[118] Since the GRADE reading comprehension test was not administered to students during the year following their participation in the program, it is not possible to directly assess whether this lack of impact on course performance during the follow-up year is due to a lack of persistence in the effect of the programs on reading comprehension.

Credit Accumulation

Table 4.2 displays the impacts of the ERO programs on students' credit attainment.[119] As discussed in Chapter 2, in order to standardize the meaning of credit accumulation across school districts, the number of core credits earned was divided by the number of core credits required for graduation by a student's district. In this way, the outcome measure is an indicator of the extent to which students have progressed toward meeting the graduation requirements in their district.

- During the program year, the ERO programs had a positive impact of 0.6 percentage point on students' accumulation of credits in core subject areas (effect size = 0.06; p-value = 0.017).[120] However, the programs did not have

---

[116]As explained in Chapter 2, the impact findings presented in this chapter are adjusted for students' background characteristics and prior achievement. As a sensitivity test, the statistical model was also specified without these covariates. "Unadjusted" impacts on course performance in the program year are similar to the "adjusted" estimates presented in this chapter. However, for the follow-up year, the unadjusted impact on GPA in core subject areas is positive and statistically significant. Further analyses show that this difference is due to the fact that there is a statistically significant difference in prior GPA between ERO and non-ERO students with school records data in the follow-up year (as reported in Chapter 2). The statistical model used in the main analysis controls for prior GPA, and should therefore remove any bias that may arise from this difference. See Appendix F for the unadjusted impact estimates.

[117]Note that the GPA measure is not cumulative and only includes students' grades in core courses during the given school year.

[118]The impact on students' ELA grades during the follow-up year is positive and statistically significant (effect size = 0.08; p-value = 0.005). However, because the impact on GPA in all core subjects is *not* statistically significant, the statistical significance of the finding for ELA grades could be due to chance (Type I error) and should be interpreted with caution.

[119]Confidence intervals for the impacts on credit accumulation in the program and follow-up year can be found in Table F.2, in Appendix F.

[120]The statistical significance of estimated impacts on GPA and credit accumulation in core courses does not change when p-values are adjusted for multiple hypothesis testing. See Appendix F for these adjusted p-values.

The Enhanced Reading Opportunities Study

Table 4.2

Impacts on Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools					
End of program year					
Credits earned in core subject areas (%)	21.4	20.9	0.6*	0.06*	0.017
English language arts (ELA)	20.1	19.5	0.6*	0.06*	0.032
Social studies	20.9	20.5	0.4	0.03	0.221
Science	23.4	22.6	0.8	0.05	0.057
Math	22.3	21.6	0.7	0.05	0.072
Sample size	2,937	2,213			
End of follow-up year (cumulative)					
Credits earned in core subject areas (%)	44.4	43.8	0.5	0.03	0.212
English language arts (ELA)	42.5	41.6	0.9	0.05	0.056
Social studies	43.2	42.7	0.5	0.02	0.415
Science	48.2	46.9	1.3	0.05	0.066
Math	45.5	45.6	-0.1	0.00	0.909
Sample size	2,542	1,894			

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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 sample sizes reported in this table are for students with school records data in the relevant year.

The cumulative number of credits earned is scaled as a percentage of the total number of credits (core or subject-specific) required for graduation in a student's district.

a statistically significant impact on credit accumulation during the follow-up year.[121]

As shown in the top panel of Table 4.2, by the end of the program year, ERO students had succeeded in attaining 21.4 percent of the core credits they need to graduate, while non-ERO students completed 20.9 percent of required core credits. Thus, the ERO programs' impact is such that ERO students accumulated 0.6 percentage point more of the core credits required for graduation (effect size = 0.06; p-value = 0.017) during the year in which they were offered enrollment in the program. [122,123] Adjusting the p-value for multiple hypothesis testing does not change the statistical significance of this finding.

As seen in Table 4.2, this impact appears to be driven by the programs' effect on credit accumulation in ELA. Specifically, relative to their counterparts, ERO students had earned 0.6 percentage point more of the ELA credits required for graduation (effect size = 0.06; p-value = 0.032). The estimated impacts on credit accumulation in the other three subject areas are also positive, but are not statistically significant. The estimated impact for social studies is 0.4 percentage point (effect size = 0.03; p-value = 0.221). The estimated impact for science is 0.8 percentage point (effect size = 0.05; p-value = 0.057). The estimated impact for math is 0.7 percentage point (effect size = 0.05; p-value = 0.072).

To put the magnitude of these findings into perspective, consider that at the end of ninth grade, students will have completed one year of a four-year program. Therefore, one way to measure whether students are on track to graduate would be to look at whether or not they have completed one-fourth of the credits required for graduation. This would mean that a student who has attained 25 percent of his or her required course credits by the end of the program year would be well positioned to graduate on time. On average, across all of the districts in the study,

---

[121] As with GPA, this pattern of results is not simply due to the fact that the sample size is smaller in the follow-up year. A sensitivity analysis shows that estimated impacts in the program year for the subset of students who have school records in both the program year and the follow-up year (longitudinal sample) are positive and statistically significant, just as they are for the full school records sample. See Appendix H for these findings.

[122] Rounding may cause slight discrepancies in calculating sums.

[123] The ERO programs' impact on credit accumulation in the program year is not simply due to an increase in the credits *attempted* by ERO students. Specifically, there was no statistically significant difference between the amount of core courses attempted by ERO and non-ERO students during the program year. This indicates that ERO and non-ERO students were attempting core courses at similar rates during the program year. Therefore, the ERO programs' impact on credit accumulation can be attributed to their effect on students' success in the courses that they attempted. See Appendix H for the estimated impact of the programs on students' core course-taking patterns (credits attempted).

students were required to successfully complete 13.95 credits in core subject areas in order to graduate (4.00 credits in ELA, 3.45 credits in social studies, 3.10 credits in science, and 3.40 credits in math).[124] Therefore, on average, students who successfully completed 25 percent of their core courses would have completed approximately 3.50 credits in core subject areas at the end of the program year.[125] As shown in Figure 4.2, had students in the ERO group not been offered the opportunity to participate in the reading class, they would have earned an average of 20.9 percent of the core credits they need to graduate (as represented by the non-ERO group), which is 4.1 percentage points short of 25 percent. Thus, the ERO programs' impact on credit accumulation (0.6 percentage point) is such that these students earned 15 percent of what they needed to be on-track to graduate using this benchmark.[126]

As shown in the bottom panel of Table 4.2, however, the ERO programs no longer had a statistically significant impact on credits earned by the end of the following school year. At the end of the follow-up year, on average, non-ERO students had accumulated 43.8 percent of the credits they need to graduate, while ERO students had accumulated 44.4 percent on average.[127] The estimated impact of 0.5 percentage point is not statistically significant, so it cannot be concluded that the programs had improved students' credit accumulation by the end of the follow-up year.[128] Since the GRADE reading comprehension test was not re-administered

---

[124]All districts required students to complete 4.0 ELA credits to graduate. In the other core subject areas, students were required to complete 3.0, 3.5 or 4.0 credits to graduate, depending on state and district policies.

[125]As seen in Appendix Table H.2, ERO students attempted an average of 3.8 core credits (1.0 in ELA, 0.9 in social studies, 1.0 in science, and 1.0 in math) during the program year. This suggests that even though some districts require students to complete only three credits in certain core subject areas, most students attempt to complete a credit in each of the core subject areas during their first year.

[126]The value of 15 percent was calculated by dividing the impact (0.6 percentage point) by the percent of credits needed to attain 25 percent of their core course credits (4.1 percentage points).

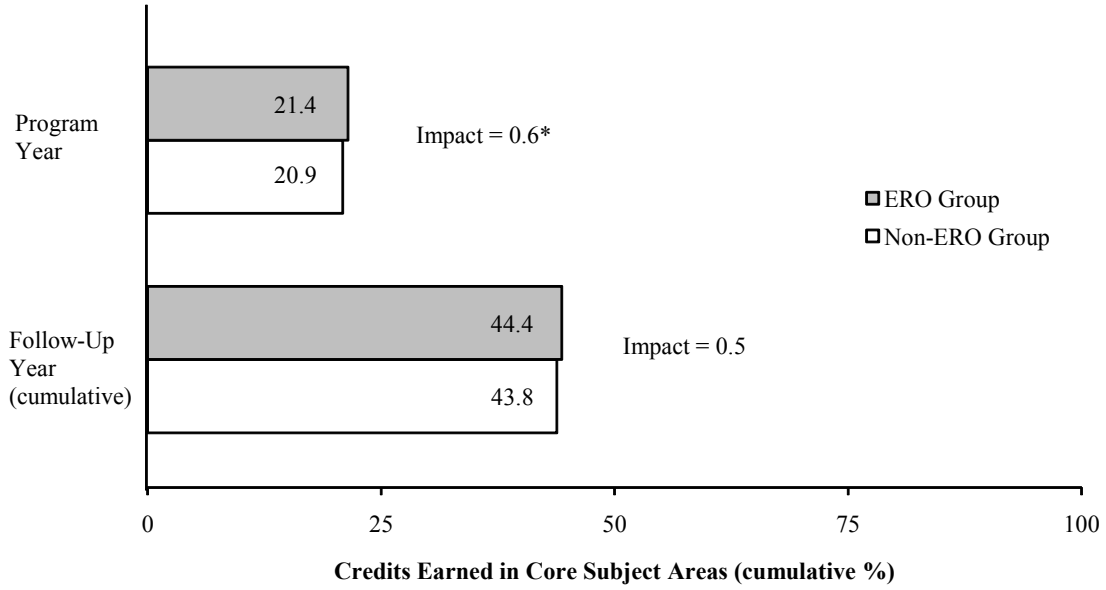
[127]Recall from Chapter 2 that the measure of credit accumulation is defined cumulatively, in order to capture a student's progress toward graduation. Thus, credit accumulation at the end of the follow-up year includes credits earned during the program year *and* the follow-up year. The estimated impact on the number of credits earned by students during the follow-up year *only* was 0.0 credits (effect size = 0.00; p-value = 0.884).

[128]Although the data are not available to test these hypotheses, it is possible that the lack of statistically significant impacts on GPA and credit accumulation at the end of the follow-up year may be explained, in part, by the ERO programs' effect on the number and difficulty of courses attempted by ERO students during the follow-up year. In the first instance, ERO students attempted 0.1 *fewer* core credits than non-ERO students during the follow-up year (effect size = -0.05; p-value = 0.032). This may contribute to explaining why by the end of the follow-up year, ERO students had no longer earned a greater number of core credits than their non-ERO counterparts. (See Appendix H for more detailed results on the ERO programs' impact on credits attempted.). Second, as discussed earlier in this section, the ERO programs had a positive impact on credits earned in the program year; therefore, in the follow-up year, students in the ERO group may have been able to attempt credits that are more advanced than the credits attempted by non-ERO students. (Or, stated otherwise, the ERO programs may have had an impact on the difficulty of courses attempted in the follow-up year.) Such an impact on course difficulty would explain, in part, why the programs no longer have an impact on GPA or credits earned by the end of the follow-up year.

The Enhanced Reading Opportunities Study

Figure 4.2

Impacts on Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample



SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment.

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 cumulative number of credits earned is scaled as a percentage of the total number of core credits required for graduation in a student's district.

during the follow-up year, it is not possible to know if the lack of impacts on credit accumulation is due to the lack of a persistent effect on reading comprehension.

State Test Scores

Another important measure of students' academic performance is their performance on standardized state tests, since it reflects students' mastery of the set of academic skills established as indicators of success by the state in which they live. Still, as discussed in Chapter 2, the timing of test administration varies by district, and in some districts the timing is also dependent on students fulfilling certain course requirements. This means that in any given year, only a subset of students in the school records sample has a state test score. In addition, the specific content of each test differs within and across districts. For example, the science assessments administered by the study districts include general science tests, as well as biology, chemistry, and physical sciences tests.[129] All of these issues complicate the interpretation of the ERO programs' impacts on state tests. For this reason, in this report, impacts on state tests are considered a secondary indicator of program effectiveness.[130] Caution should also be taken when interpreting any individual state test score impacts because of the number of hypothesis tests conducted. There are eight tests (four subject areas, for each of the two study years).

Table 4.3 presents the estimated impact of the ERO programs on state test scores in each of the four core subject areas (ELA, social studies, science, and math) during the program year and the follow-up year. Recall from Chapter 2 that test scores were converted to z-scores for the purposes of standardizing their scale. Hence, the estimated impact of the ERO programs on state test scores is expressed in terms of standard deviation units or effect size units only.[131]

- Together, the programs had a positive impact on students' performance on state tests in ELA (effect size = 0.11; p-value = 0.003) and mathematics (effect

[129]As a sensitivity test, the estimated impact of the programs on state tests was re-estimated based on a model that includes controls for the specific type of content tested (for example, biology, chemistry, or general science). The magnitudes of the estimated impacts based on this model are similar to those reported in this chapter; patterns of statistical significance are the same.

[130]As discussed in Chapter 2, since these findings are considered secondary, no multiple hypothesis testing was used. These findings are used to add contextual information to the primary impact findings.

[131]Test scores for each assessment were standardized by district, school year, and cohort, based on the mean and standard deviation of the non-ERO group.

The Enhanced Reading Opportunities Study

Table 4.3

Impacts on State Test Scores (Standardized), School Records Sample

Outcome	Number of Students	ERO Group	Non-ERO Group	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools					
Program year					
English language arts (ELA)	2,244	0.11	0.01	0.11*	0.003
Social studies	952	0.07	0.01	0.06	0.343
Science	2,348	0.07	0.01	0.06	0.166
Math	2,668	0.08	0.01	0.07*	0.045
Follow-up year					
English language arts (ELA)	2,408	0.00	-0.01	0.01	0.705
Social studies	2,237	-0.04	0.01	-0.04	0.262
Science	2,661	0.00	0.02	-0.02	0.649
Math	2,537	-0.03	0.02	-0.04	0.265

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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 numbers of students reported in this table are for students in the school records sample who have state test scores for a given subject area in the relevant year. A student may have taken more than one test in a given subject area in more than one year. If a student wrote a specific test more than once, only his or her first score is used.

Rounding may cause slight discrepancies in calculating sums and differences.

State test scores are standardized by district, follow-up year, and cohort, using the means and standard deviation of the non-ERO group.

size = 0.07; p-value = 0.045) in the program year.[132] There are no statistically significant impacts on state test scores in the follow-up year.

The ERO programs' impact on ELA test scores is consistent with their positive impact on students' GRADE reading comprehension scores in the program year (see Chapter 3) as well as their impact on students' performance in ELA courses in the program year (reported in the previous section). Hence, in the program year, ERO students were able to achieve higher grades and earn more credits in ELA, and achieve higher scores on ELA state tests than their non-ERO counterparts. The associations between these outcomes are explored in more detail in Chapter 5.

Student Behaviors

While the primary focus of the reading programs is on the teaching of reading strategies, both interventions also focus on teaching positive behavioral norms. Table 4.4 presents the estimated impact of the two programs on student's attendance rate (as a percentage of the number of days that a student is enrolled), and whether a student was suspended during the school year.[133]

- Together, the ERO programs did not affect students' school behaviors in either the program year or the follow-up year.

As shown in the top panel of Table 4.4, on average, ERO students attended 93.5 percent of school days during the program year, while non-ERO students attended 93.2 percent of school days during that year. The estimated impact of 0.3 percent is not statistically significant (effect size = 0.03; p-value = 0.144). Also during the program year, 32.0 percent of ERO students were suspended at least once during the school year, while 33.1 percent of non-ERO students were ever suspended during that year. The difference of -1.2 percent is not statistically

[132]To further explore the magnitude of these positive impacts, the ERO study team conducted an analysis measuring the effects of the programs on a dichotomous variable indicating whether or not students met their states' proficiency standards in these two subject areas and found that 4.2 percent more ERO students met state proficiency standards on ELA tests taken during the program year than non-ERO students. This estimated impact was statistically significant (effect size = 0.08; p-value = 0.020). The estimated impact on meeting state proficiency standards on math tests during the program year was not statistically significant. See Appendix H for detailed findings.

[133]This measure includes in-school suspensions, out-of-school suspensions, and expulsions. Five of the 34 schools were not included in this analysis due to missing discipline data. One district (four schools) provided information on students' disciplinary infractions, but not on the disciplinary actions then taken by the district. For these schools, students were counted as being suspended if they committed at least one serious infraction (such as fighting), or were disciplined for five or more less serious infractions (such as disruptive behavior).

The Enhanced Reading Opportunities Study

Table 4.4

**Impacts on Student Behaviors,
School Records Sample**

Outcome	Number of Students	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools						
Program year						
Attendance rate (%)	5,101	93.5	93.2	0.3	0.03	0.144
Ever suspended (%)	4,597	32.0	33.1	-1.2	-0.02	0.365
Follow-up year						
Attendance rate (%)	4,348	91.2	90.6	0.6	0.04	0.141
Ever suspended (%)	3,936	32.2	33.9	-1.7	-0.04	0.222

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The “Non-ERO Group” values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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 numbers of students reported in this table are for students in the school records sample and with data on the specific student behavior outcomes in the relevant year. There are five schools with missing data for "ever suspended."

Rounding may cause slight discrepancies in calculating sums and differences.

The attendance rate is defined relative to the number of days that the student was enrolled in the district.

"Ever suspended" measures the percentage of students suspended (in school or out of school) or expelled one or more times during the relevant school year.

significant (effect size = -0.02; p-value = 0.365). Similar to the program year, estimated impacts in the follow-up year are not statistically significant.

Impacts, by Student Subgroup

As in the first two reports, subgroup analyses were conducted to compare the impacts of the programs on several subgroups of students, including:

- Students with different reading comprehension performance scores at baseline (2.0-3.0 years below grade level, 3.1-4.0 years below grade level, and 4.1-5.0 years below grade level)
- Students who were overage for grade at the start of the program and those who were not overage for grade
- Students from multilingual families and students from English-only families

Although the estimated impact of the ERO programs on high school outcomes is statistically significant for some student subgroups, in these instances the difference in impacts between subgroups is not statistically significant.[134] Therefore, the findings from these analyses do not offer conclusive evidence that the programs were more effective in improving the school-based outcomes of any of these groups of students over others. Estimated impacts by subgroup are reported in Appendix H.

Impacts, by Program

This section presents impact findings for each of the two reading programs separately (17 Reading Apprenticeship schools and 17 Xtreme Reading Schools).[135] Since the focus of the study is on the type of intervention represented by both the programs combined, this section should be viewed primarily as a way to better understand the pooled findings. Program-specific impacts on students' course performance tell a similar story as the pooled findings:

- Both programs had a positive impact on students' GPA in core subject areas during the program year. The estimated impact for RAAL schools is 0.07

[134]Among the three subgroups of students defined by their baseline reading comprehension levels, there is a statistically significant difference in the impact of the ERO programs on credit accumulation in the follow-up year; however, the estimated impact on this outcome is not statistically significant for any of the three subgroups individually.

[135]See Appendix A for a detailed summary of the similarities and differences between the two programs.

point and the estimated impact for Xtreme Reading schools is 0.06 point. Neither program had a statistically significant effect on students' core GPA in the follow-up year.

- Although neither program had a statistically significant impact on credits earned in core subject areas during the program year, the magnitude of each program's impact estimate (0.6 percentage point for RAAL and 0.5 percentage point for Xtreme Reading) are similar to that of the overall impact estimate. Program-specific impacts in the follow-up year were not statistically significant.

- It cannot be concluded that one program was more effective than the other. Differences in impacts between the two programs on the two key indicators of academic performance — core GPA and credit accumulation — are not statistically significant in either year.

As explained in Chapter 1, the ERO study was designed primarily to test the combined impact of RAAL and Xtreme Reading as a broad class of reading intervention. The study's design also makes it possible to evaluate the effectiveness of each of these two programs separately as is done in this section, but the study design only allows for the detection of relatively large differences in impacts between the programs. For this reason, it is not possible to offer conclusive evidence about the difference in impacts between the two programs and any discussion of the program differences should be considered informational.

Course Performance

For the two primary indicators of program effectiveness — GPA and credit accumulation — the results by program are similar to the findings for both programs pooled together. In addition, estimated impacts on these outcomes are not statistically different between the two programs. For these reasons, it cannot be concluded that impacts on course performance differed by program.

GPA

Table 4.5 presents impacts on students' GPA in core subject areas for each reading program separately.[136] Both RAAL and Xtreme Reading had a positive impact on students' GPA in

[136]Confidence intervals for the impacts on GPA and credit accumulation by program for students in the program and follow-up year can be found in Table F.3, in Appendix F.

The Enhanced Reading Opportunities Study

Table 4.5

Impacts on Grade Point Average (GPA), School Records Sample, by Program

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
Reading Apprenticeship schools					
Program year					
GPA in core subject areas	1.54	1.47	0.07*	0.07*	0.019
English language arts (ELA)	1.62	1.54	0.07	0.06	0.069
Social studies	1.65	1.56	0.09*	0.08*	0.029
Science	1.49	1.35	0.13*	0.12*	0.001
Math	1.44	1.43	0.02	0.01	0.696
Sample size	1,468	1,095			
Follow-up year					
GPA in core subject areas	1.55	1.52	0.03	0.03	0.392
English language arts (ELA)	1.72	1.67	0.05	0.04	0.263
Social studies	1.63	1.57	0.06	0.05	0.203
Science	1.49	1.44	0.05	0.05	0.267
Math	1.37	1.40	-0.03	-0.03	0.531
Sample size	1,285	927			
Xtreme Reading schools					
Program year					
GPA in core subject areas	1.66	1.60	0.06*	0.06*	0.038
English language arts (ELA)	1.80	1.73	0.07	0.06	0.068
Social studies	1.68	1.59	0.08*	0.07*	0.041
Science	1.59	1.56	0.03	0.03	0.460
Math	1.56	1.50	0.06	0.05	0.124
Sample size	1,469	1,118			
Follow-up year					
GPA in core subject areas	1.62	1.56	0.06	0.06	0.068
English language arts (ELA)	1.82	1.70	0.13*	0.11*	0.004
Social studies	1.63	1.63	0.00	0.00	0.948
Science	1.55	1.49	0.06	0.05	0.167
Math	1.51	1.47	0.05	0.04	0.284
Sample size	1,257	967			

(continued)

Table 4.5 (continued)

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The sample sizes reported in this table are for students with school records data in the relevant year. Because students may not have earned grades in all core subject areas in a given year, sample sizes differ for impacts in the specific core subject areas.

Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

core subject areas during the program year. The estimated impact for RAAL schools is 0.07 point (effect size = 0.07; p-value = 0.019), while the effect size for Xtreme Reading schools is 0.06 point (effect size = 0.06; p-value = 0.038). The subject-specific findings in Table 4.5 show that in RAAL schools, these impacts are driven by the program's effect on students' GPA in social studies where there is an estimated impact of 0.09 point (effect size = 0.08; p-value = 0.029) and science where there is an estimated impact of 0.13 point (effect size = 0.12; p-value = 0.001). In Xtreme Reading schools, the programs have a positive estimated impact of 0.08 point on students' grades in social studies (effect size = 0.07; p-value = 0.041). However, like the pooled results, neither program had a statistically significant impact on GPA in core subject areas during the follow-up year. Differences in impacts between the two programs are not statistically significant in either the program year or the follow-up year.

Credit Accumulation

Table 4.6 displays the impacts of RAAL and Xtreme Reading on students' credit accumulation. Neither of the programs had a statistically significant impact on credits earned in core subject areas during the program year. The estimated impact is 0.6 percentage point (effect size = 0.06; p-value = 0.072) for RAAL schools; for the Xtreme Reading schools, the estimated impact is 0.5 percentage point (effect size = 0.05; p-value = 0.138). Still, the magnitude of the

The Enhanced Reading Opportunities Study

Table 4.6

Impacts on Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample, by Program

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
Reading Apprenticeship schools					
End of program year					
Credits earned in core subject areas (%)	20.5	19.9	0.6	0.06	0.072
English language arts (ELA)	18.8	18.1	0.7	0.06	0.092
Social studies	20.6	20.6	0.1	0.01	0.881
Science	22.0	21.2	0.9	0.05	0.136
Math	21.9	20.7	1.1*	0.08*	0.034
Sample size	1,468	1,095			
End of follow-up year (cumulative)					
Credits earned in core subject areas (%)	42.6	42.1	0.4	0.03	0.486
English language arts (ELA)	40.2	39.5	0.7	0.04	0.322
Social studies	42.3	42.5	-0.2	-0.01	0.813
Science	45.8	44.6	1.2	0.05	0.229
Math	43.9	43.5	0.3	0.01	0.717
Sample size	1,285	927			
Xtreme Reading schools					
End of program year					
Credits earned in core subject areas (%)	22.3	21.8	0.5	0.05	0.138
English language arts (ELA)	21.5	21.0	0.5	0.04	0.213
Social studies	21.2	20.4	0.8	0.05	0.105
Science	24.8	24.1	0.7	0.04	0.237
Math	22.8	22.6	0.2	0.01	0.702
Sample size	1,469	1,118			
End of follow-up year (cumulative)					
Credits earned in core subject areas (%)	46.2	45.5	0.7	0.04	0.266
English language arts (ELA)	44.8	43.6	1.1	0.06	0.084
Social studies	44.1	42.9	1.2	0.05	0.138
Science	50.6	49.1	1.5	0.06	0.143
Math	47.2	47.7	-0.5	-0.02	0.590
Sample size	1,257	967			

(continued)

Table 4.6 (continued)

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The sample sizes reported in this table are for students with school records data in the relevant year.

The cumulative number of credits earned is scaled as a percentage of the total number of credits (core or subject-specific) required for graduation in a student's district.

impacts for each program are similar to the magnitude of the estimated impacts for both programs pooled together (effect size = 0.06; p-value = 0.017) and they are not statistically different from each other.[137] Similar to the pooled sample, there are no statistically significant impacts on cumulative credits earned for either program during the follow-up year.

State Test Scores

Table 4.7 presents impacts on state test scores for each of the two reading programs. As with the pooled impacts, caution should be taken when interpreting these findings due to the inconsistencies in subject matter and timing of test-taking within and across school districts. As shown in Table 4.7, the RAAL program had a positive impact on students' ELA test scores during the program year (effect size = 0.15; p-value = 0.006). Although the estimated impact on this outcome is not statistically significant for Xtreme Reading schools, the difference in impacts between the two programs is not statistically significant, and therefore it cannot be con-

[137]Since, the sample size for each of the programs is approximately half the sample size of the full sample, the magnitude of the impact needed for statistical significance is larger for the programs than the full sample.

The Enhanced Reading Opportunities Study

Table 4.7

Impacts on State Test Scores, School Records Sample, by Program

Outcome	Number of Students	ERO Group	Non-ERO Group	Estimated Impact Effect Size[a]	P-Value for Estimated Impact
Reading Apprenticeship schools					
Program year					
English language arts (ELA)	1,053	0.13	-0.02	0.15*	0.006
Social studies	447	0.10	-0.06	0.16	0.088
Science	1,151	0.05	-0.10	0.14*	0.033†
Math	1,263	0.09	0.04	0.05	0.320
Follow-up year					
English language arts (ELA)	1,140	0.00	-0.03	0.03	0.625
Social studies	1,089	0.05	-0.01	0.06	0.285†
Science	1,269	0.06	0.07	-0.01	0.822
Math	1,226	-0.04	-0.01	-0.03	0.564
Xtreme Reading schools					
Program year					
English language arts (ELA)	1,191	0.10	0.03	0.08	0.117
Social studies	505	0.05	0.06	-0.01	0.873
Science	1,197	0.10	0.13	-0.03	0.588
Math	1,405	0.07	-0.01	0.08	0.076
Follow-up year					
English language arts (ELA)	1,268	0.00	0.00	0.00	0.965
Social studies	1,148	-0.12	0.01	-0.13*	0.017
Science	1,392	-0.06	-0.04	-0.01	0.789
Math	1,311	-0.02	0.04	-0.06	0.312

(continued)

Table 4.7 (continued)

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The number of students reported in this table are for students in the school records sample who have state test scores for a given subject area in the relevant year. If a student wrote a specific test more than once, only the first score is used.

Rounding may cause slight discrepancies in calculating sums and differences.

[a]State test scores are standardized by district, follow-up year, and cohort, using the mean and standard deviation of the non-ERO group in the relevant year.

cluded that RAAL had a different effect on students' ELA state test scores than Xtreme Reading.

The RAAL program also had a positive impact on students' science test scores during the program year, with an effect size of 0.14 (p-value = 0.033). The estimated impact on science test scores for students in Xtreme Reading schools was not statistically significant (effect size = -0.03; p-value = 0.588). The difference in impacts between the two programs is statistically significant (p-value = 0.044), suggesting that impacts on science tests is greater in RAAL schools than in Xtreme Reading schools during the program year. However, because state tests are available only in a subset of the study schools and because they are a secondary outcome in this analysis, this finding does not provide strong evidence of a differential impact on science test scores and could be due to chance.[138] Also, as shown in Table 4.3, there is no impact on

[138]See Chapter 2 for a discussion of multiple hypothesis testing.

science state test scores for both programs combined, which means that the statistical significance of the program-specific impact of RAAL on science test scores may also be due to chance.

Neither program had a positive impact on state test scores during the follow-up year. In fact, the Xtreme Reading program had a negative impact on social studies test scores (effect size = -0.13; p-value = 0.017). This negative impact is statistically different (p-value = 0.015) from the impact of the RAAL program (effect size = 0.06; p-value = 0.285).

Student Behaviors

Table 4.8 displays the impacts of the two programs separately on student attendance and whether a student was ever suspended. As with the pooled sample, estimated impacts on these measures are not statistically significant during the program year. However, in RAAL schools during the follow-up year, 4.3 percent fewer ERO students were suspended compared with non-ERO students (effect size = -0.09; p-value = 0.025). But, this impact is not statistically different from the impact on this outcome in Xtreme Reading schools, therefore it cannot be concluded that the impacts on suspensions differ by program. As discussed earlier, the two programs combined did not have a statistically significant impact on whether students were suspended during the follow-up year.

In summary, this chapter establishes that the ERO programs had a positive impact on students' performance in their course work and on some state tests during the year in which the students were offered the opportunity to participate in the programs, but not in the following year. This pattern of impacts is explored in the next chapter.

The Enhanced Reading Opportunities Study

Table 4.8

Impacts on Student Behaviors, School Records Sample, by Program

Outcome	Number of Students	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
Reading Apprenticeship schools						
Program year						
Attendance rate (%)	2,528	92.9	92.3	0.6	0.06	0.089
Ever suspended (%)	2,346	30.8	32.4	-1.6	-0.03	0.370
Follow-up year						
Attendance rate (%)	2,160	90.5	90.2	0.3	0.02	0.545
Ever suspended (%)	2,016	31.7	35.9	-4.3*	-0.09*	0.025
Xtreme Reading schools						
Program year						
Attendance rate (%)	2,573	94.1	94.0	0.0	0.00	0.880
Ever suspended (%)	2,251	33.2	34.0	-0.8	-0.02	0.665
Follow-up year						
Attendance rate (%)	2,188	91.9	91.1	0.8	0.06	0.133
Ever suspended (%)	1,920	32.7	31.9	0.8	0.02	0.673

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The numbers of students reported in this table are for students in the school records sample and with data on the specific student behavior outcomes in the relevant year. There are five schools with missing data for "ever suspended."

Rounding may cause slight discrepancies in calculating sums and differences.

The attendance rate is defined relative to the number of days that the student was enrolled in the district.

"Ever suspended or expelled" measures the percentage of students suspended (in school or out of school) or expelled one or more times during the relevant school year.

Chapter 5

Exploring Hypotheses About Program Effects and Their Persistence

Overall, the findings from this study show that the Enhanced Reading Opportunities (ERO) programs had a positive impact on an array of academic outcomes during the year in which students were enrolled in the supplemental reading program — with effect sizes that range from 0.06 (core credits earned) to 0.11 (state English language arts test scores) and including an impact on reading comprehension of 0.09 or two additional months of learning. However, the ERO programs did not improve student outcomes in the following school year, when students were no longer receiving the literacy supports provided by the programs.

In order to both better understand this pattern of findings and inform future research on adolescent literacy, in this chapter hypotheses are explored about aspects of program design and delivery that might influence the magnitude of the ERO programs' effect on student achievement in ninth grade, as well as the persistence of these effects after students have completed the program. These hypotheses are examined by exploiting the fact that there is variation in the impact of the program across the study schools. It is important to note that these analyses are nonexperimental and may not reflect true causal relationships between implementation and impacts.

The Relationship Between Implementation Characteristics and Program Impacts

The relationship between school-level impacts and the characteristics of implementation is examined in two stages. The first stage provides an assessment of overall variation in impacts on academic performance across the 34 participating schools. To the extent that there is variation in impacts across the schools, 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 the question of whether program fidelity and a school's overall focus on literacy could be related to impacts on student achievement and reading behavior outcomes.

Variation in Impacts Across Schools

In order to examine the relationship between implementation characteristics and impacts, the first step is to examine the extent to which impacts differ across the schools in the

study. Of particular interest is whether (1) the ERO programs had a larger impact on ninth-grade achievement outcomes in some schools than others and (2) whether the programs had a positive impact on tenth-grade academic outcomes in some schools, even though on average the estimated impact of the program was not statistically significant in the follow-up year. For simplicity, the findings in this section focus on the ERO programs' impact on students' grade point average (GPA) in core subject areas. Similar analyses for credit accumulation can be found in Appendix K.[149]

- Figures 5.1 and 5.2 present the estimated impact of the ERO programs on GPA, in the program year and in the follow-up year, respectively, for each of the 34 schools in the study. These figures also show the 95 percent confidence interval around each impact estimate; 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). As seen in these figures: In the program year (ninth grade), estimated impacts on GPA range from -0.23 point to 0.34 point across the study schools (compared with the average impact of 0.06 point). Four of the estimates are both positive and statistically significant.

- In the follow-up year (tenth grade for most students), estimated impacts on GPA range from -0.57 point to 0.14 point across the study schools (compared with the average impact of 0.05 point). Two of the estimates are positive and statistically significant.

Both figures also show that schools implementing a given reading program (RAAL, Xtreme Reading) are located throughout the range of impacts. This supports the previous chapter's finding that on average, the two programs' estimated impact on GPA does not differ by a statistically significant amount, whether in the program year or the follow-up year.

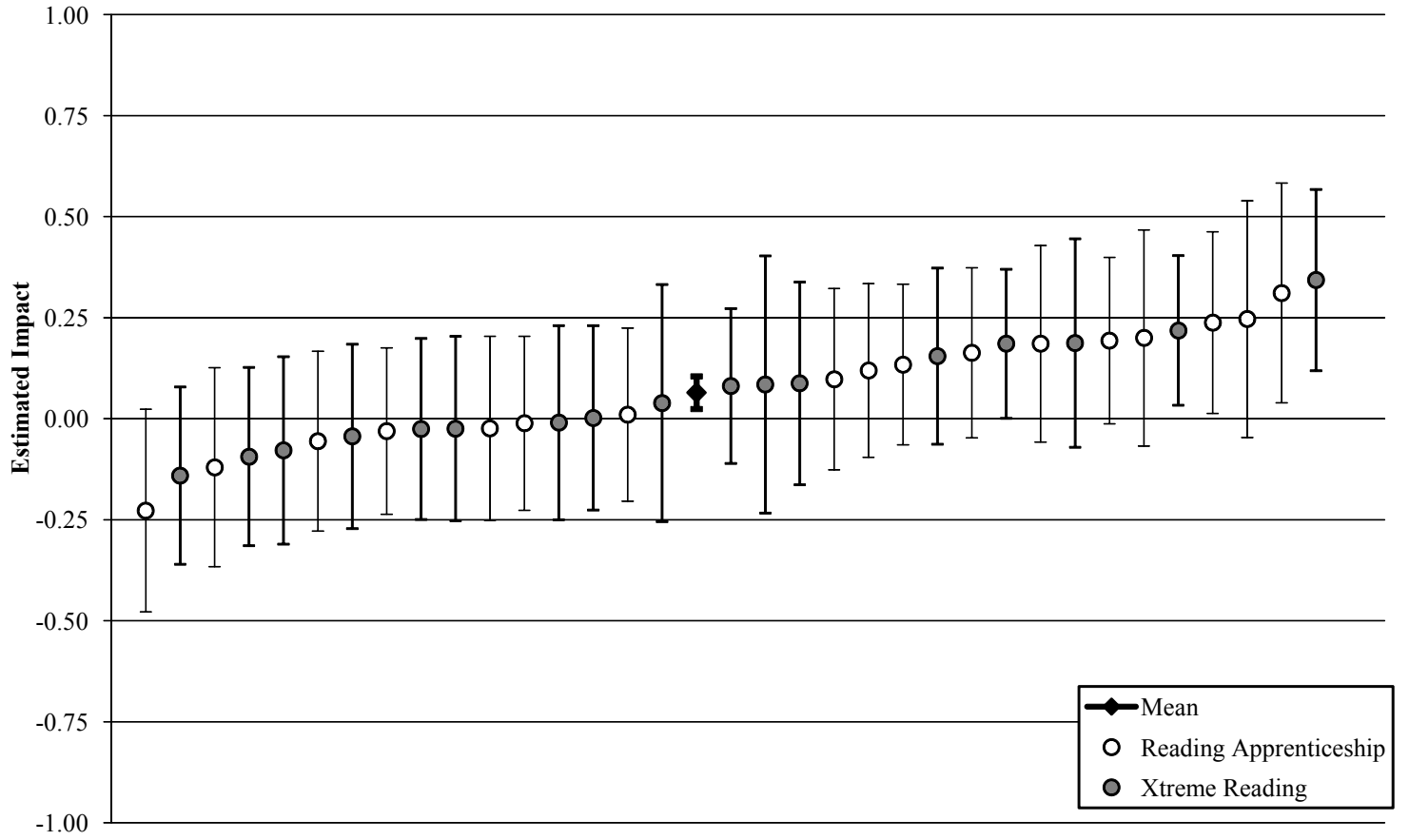
Because impacts are estimated with error, however, the variation in estimated impacts displayed in Figures 5.1 and 5.2 overstates the true variation in program impacts across schools. To assess the variability in impacts across schools more systematically, a composite F-test was used to assess whether the variation in school-level impacts is larger than would be expected

[149]Impacts on state tests are not examined in this analysis because state test scores are available only for a subset of the study schools in any given follow-up year (and thus, there is less variation in impacts to explore).

The Enhanced Reading Opportunities Study

Figure 5.1

Fixed-Effect Impact Estimates on Grade Point Average (GPA) in Core Subject Areas in the Program Year, by School, School Records Sample



(continued)

Figure 5.1 (continued)

SOURCE: MDRC calculations from individual students' school records data supplied by each school district.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2.

The fixed-effects impact estimates in this figure are the regression-adjusted coefficients on the interaction between schools and the treatment indicator. These estimates are adjusted for random baseline differences between the ERO and the non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation.

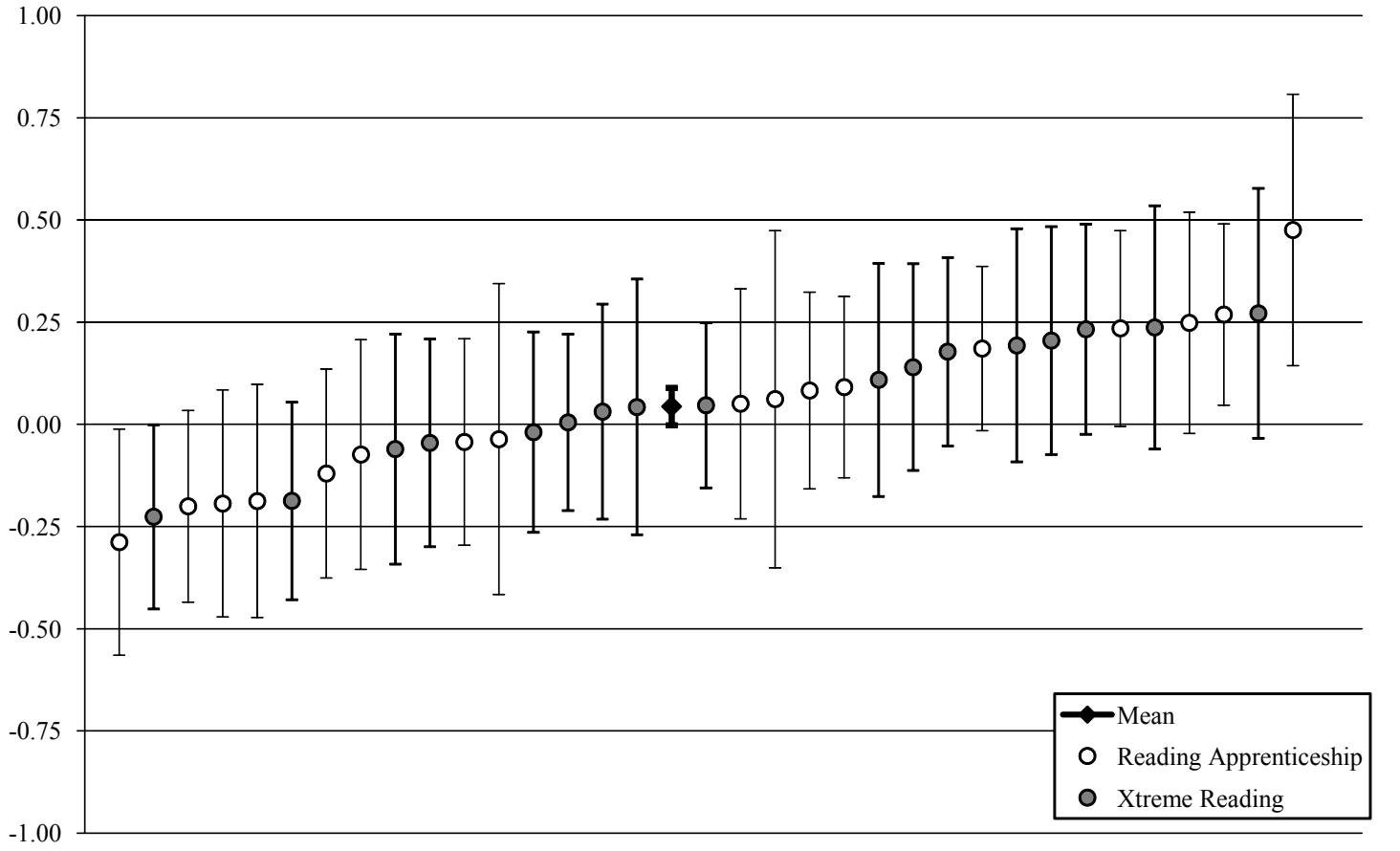
GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

Sample size: 5,073 students.

The Enhanced Reading Opportunities Study

Figure 5.2

Fixed-Effect Impact Estimates on Grade Point Average (GPA) in Core Subject Areas in the Follow-Up Year, by School, School Records Sample



(continued)

Figure 5.2 (continued)

SOURCE: MDRC calculations from individual students' school records data supplied by each school district.

NOTES: The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The fixed-effects impact estimates in this figure are the regression-adjusted coefficients on the interaction between schools and the treatment indicator. These estimates are adjusted for random baseline differences between the ERO and the non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

Sample size: 4,359 students.

due to chance.[150] The results from these tests show that school-to-school variation in impacts is statistically significant in the follow-up year. Therefore, even though the ERO programs did not improve GPA in the tenth grade on average, the programs had a larger impact on tenth-grade academic performance in some of the study schools than others.[151] The next section examines the conditions under which the ERO programs' impact may be larger in the program year and more sustained in the follow-up year.[152]

Association Between Impacts and Implementation Characteristics

This section examines the association between the ERO programs' impact in the study schools (as shown in Figures 5.1 and 5.2) and the implementation characteristics of these schools, which carries practical relevance for policymakers and practitioners. The analysis focuses on two types of implementation features that could potentially influence the size and persistence of program impacts:

- **Implementation fidelity:** It is hypothesized that impacts in the program year and the follow-up year may be larger when the supplemental programs are better aligned to developers' specifications for the program. In this study, implementation fidelity ratings are available from classroom observations and described in Chapter 3.[153]

- **Other literacy support services:** It is hypothesized that the ERO programs may have a larger impact on student outcomes in the follow-up year (tenth grade) if students have access to other literacy support services after they leave the program.[154] The ERO student survey — which was administered at

---

[150]See Appendix K for details on the statistical model used to estimate impacts and to test whether the variation in impacts is statistically significant.

[151]The *difference* between impacts in the program year and the follow-up year also differs by a statistically significant amount across the study schools (p-value = 0.001). This suggests that in the follow-up year, the estimated impact of the program decreased by a smaller amount in some of the study schools than in others.

[152]In order to estimate the amount of true variation in the impacts across the study schools, the school-to-school variation in estimated program impacts can be corrected for noise. The "noise-corrected" standard deviation in impacts across schools is 0.06 point in the program year and 0.10 point in the follow-up year. Appendix K provides further detail on how the noise-corrected variation in impacts was obtained.

[153]The ratings are based on a scale of 1 to 3. Appendix C provides further detail on the measurement of implementation fidelity in this study.

[154]In a qualitative study of another supplemental reading program, for example, interviewed students reported that after leaving the program, they no longer used the reading strategies that they had learned (Diamond, Corrin, and Levinson, 2004). Given these findings, students may benefit from a more "transitional" structure in which they are provided literacy supports in the following school year to help them continue to apply the reading strategies in their regular course work.

the end of ninth grade — includes items on the number of sessions of extra literacy help that students received in classes at the school and from an adult tutor at the school.[155] In this analysis, the average amount of such services received by non-ERO students during their ninth-grade year (by school and cohort) is used as a proxy for a school's general focus on literacy services across ninth and tenth grade.[156]

Table 5.1 presents the estimated association between these implementation characteristics and the ERO programs' impact on GPA in core subject areas in the program year and the follow-up year. The values in this table are standardized regression coefficients. That is, they represent the change in the impact on core GPA that is predicted by a one standard deviation increase in the implementation characteristic; this change in impact is scaled as an effect size (that is, as a proportion of the standard deviation in GPA).[157,158] When interpreting the results in this table, it is important to remember that associations between impacts and implementation characteristics do not necessarily represent true causal relationships, since these characteristics were not randomly assigned to the study schools.

As seen in Table 5.1, the implementation fidelity ratings are not associated with impacts on GPA, either in the program year or the follow-up year. That is, there is no evidence to support the hypothesis that more faithful implementation of the programs is related to larger effects on student outcomes. Nor is there evidence that the intensity of available literacy supports in school is associated with impacts in the program year.

However, there is a statistically significant relationship between impacts on GPA in core subject areas in the follow-up year (tenth grade) and the extent to which schools provide students with adult literacy tutoring in the program year (p-value = 0.026). The magnitude of this association is such that 8 extra tutoring sessions during the school year (the standard deviation of this measure) is associated with an increase of 0.07 in the ERO programs' impact in the follow-up year (in effect size), which is equal to the size of the estimated impact of the

---

[155]Appendix B provides further detail on these measures of supplemental literacy services.

[156]This assumes that schools where students receive more support services on average in ninth grade also receive more support services in tenth grade.

[157]Regression coefficients were obtained by fitting an impact model that includes an indicator of treatment status, as well as a set of interactions between the treatment indicator and the implementation conditions; the values reported in the table are the coefficients on these interaction terms. All models also control for the blocking of random assignment by school and cohort, as well as students' baseline characteristics and prior achievement. Appendix K provides further detail on the statistical model used to conduct the analysis.

[158]See Appendix K for associations between these implementation characteristics and impacts on credit accumulation.

The Enhanced Reading Opportunities Study

Table 5.1

Association Between Program Implementation Conditions and Impacts on Grade Point Average (GPA) in Core Subject Areas (Standardized Regression Coefficients)

Implementation Conditions (Program Year)	Association with Impacts on GPA in the: Program Year	Follow-Up Year
ERO program fidelity[a]		
Composite implementation fidelity rating		
Estimated coefficient	0.00	-0.01
Standard error	(0.02)	(0.03)
P-value	0.828	0.563
Non-ERO supplemental literacy support services[b]		
Number of school-based literacy classes		
Estimated coefficient	-0.02	-0.02
Standard error	(0.03)	(0.03)
P-value	0.350	0.521
Number of tutoring sessions with school-based adult tutor		
Estimated coefficient	0.03	0.07*
Standard error	(0.03)	(0.03)
P-value	0.227	0.026
Sample size	5,150	4,436
R-squared	0.423	0.360

SOURCES: MDRC calculations from the Enhanced Reading Opportunities school records data, classroom observations, and follow-up survey.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The values in this table are standardized regression coefficients that measure the extent to which the implementation conditions in the study schools are associated with impacts on academic performance in the program year and follow-up year. Entries in the table represent the estimated association between a 1 standard deviation increase in the implementation measure and the magnitude of the impact of the ERO programs on academic performance; these associations are scaled as a proportion of the standard deviation in the academic performance measure for the non-ERO group (i.e., as an effect size).

Regression coefficients were obtained by fitting an impact model that includes an indicator of treatment status, as well as a set of interactions between the treatment indicator and the implementation conditions; the values reported in this table are the coefficients on these interaction terms. All models also control for the blocking of random assignment by school and cohort, as well as the following covariates: whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by their school district in the year prior to ERO participation, a baseline measurement of the outcome variable in the school year prior to ERO participation, and indicators of missing data for all relevant student characteristics.

Statistical significance of the coefficients is indicated (*) when the p-value is less than or equal to 5 percent.

[a]Fidelity ratings are based on classroom observations conducted in the spring of the program year (spring 2005 for Cohort 1 and spring 2006 for Cohort 2). The composite rating captures implementation fidelity on two key dimensions: the classroom learning environment and the teacher's use of instructional strategies focused on reading comprehension. The composite rating is on a scale of 1 to 3.

[b]These measures are from the follow-up student survey, administered at the end of the program year (spring 2005 for Cohort 1 and spring 2006 for Cohort 2). They are based on the average number of literacy classes or tutoring sessions taken by students in the non-ERO group in each school and cohort during the program year.

programs on GPA in the program year.[159] One possible interpretation of this finding is that impacts in tenth grade may be larger (and more sustained) in schools where literacy tutoring is available to ERO students in the follow-up year, because this type of support may help students "transition" out of the program and continue to use the ERO reading strategies in their tenth-grade course work. Because these findings are nonexperimental, however, the association between impacts and tutoring services may actually be explained by some other school characteristic that is not measured in the analysis.

[159]Note that in the program year (ninth grade), one would expect the association between literacy supports and program impacts to be the zero or negative, because a more literacy-rich environment would reduce the contrast between the reading services received by the ERO group and the non-ERO group. The results in Table 5.1 support this expectation.

Chapter 6

Program Costs and Poststudy Implementation

The earlier chapters in this report have shown that the Enhanced Reading Opportunities (ERO) programs improve students' outcomes while they are enrolled in the programs, but that these impacts do not persist afterwards. When deciding whether and how to pursue an initiative such as this one, it is useful for educational practitioners to understand the programs' costs (as implemented in the study and in a nonstudy setting) and the extent to which other educators valued the programs enough to sustain them. Accordingly, this chapter presents findings on these two aspects of the ERO interventions, which are intended to be of practical use to policymakers and educators who are considering adopting an intensive, full-year intervention for struggling adolescent readers.

The chapter begins by presenting information on the costs associated with providing the ERO programs in the context of this study. The costs of the reading programs are disaggregated by type of resource (such as salaries, training, travel, and equipment), so that education practitioners who are considering this type of intervention can better estimate what it would cost to implement the ERO programs in their own schools. The key findings of this cost analysis are the following:

- As implemented in the study, the average cost of the one-year ERO programs is $1,931 per student. Costs were higher in the first implementation year ($2,035 per student) than in the second year ($1,829 per student) due to startup costs in the first year (initial training, materials, and equipment).

- Salary expenditures represent the largest portion of programs' costs, representing 72 percent of the average per-student cost. This means that in a nonstudy setting, the per-student cost of implementing the programs would depend primarily on the educational attainment and years of service of the teachers assigned to teach the program.

Next, the chapter examines whether schools continued to use the ERO programs beyond the two-year implementation period required by the study. After the study ended, ERO schools and districts continued to receive Smaller Learning Communities (SLC) grant funding, but they were free to decide whether to keep implementing the ERO programs or to use the funds to improve other aspects of their SLCs. Whether schools in the study continued to implement the ERO programs during this period provides a measure of the longer-term appeal of these programs. The chapter also examines in what ways (if any) schools that continued to use the ERO programs modified them to fit local circumstances; however, there is no way to know whether

any poststudy divergence from the program models would lead to the same impacts as those estimated during the study period. These findings are based on interviews with staff at 30 of the 34 ERO high schools. The key findings from this investigation are the following:

- Fourteen high schools (47 percent of interviewed schools) continued to offer the ERO program after the end of the study-required implementation period, though they modified some aspects of the programs' implementation. These modifications were typically aimed at increasing the number of students served (increasing class size limits, expanding eligibility criteria) and/or lengthening the duration of the programs to serve students in more than just ninth grade.

The chapter then concludes by presenting estimates of what it would cost to implement the ERO programs in a "nonstudy" setting. These estimates are based on assumptions that are guided by the study team's findings about how schools implemented and modified the programs after the end of the study.

The Cost of Implementing the ERO Programs

This section discusses the costs associated with providing the ERO programs as they were implemented in the ERO study. The cost of the ERO programs is presented for each type of resource needed to implement the ERO programs, which are:

- **Personnel salaries and associated fringe benefits:** Salaries and benefits for the following staff:
  - One full-time equivalent (FTE) ERO teacher per school
  - One district-level ERO program coordinator (estimated at 8.75 percent FTE per school)[160]
- **Training costs:** Costs associated with the summer training institutes, midyear "booster" training sessions, and visits from coaches, including:[161]

---

[160]Each district assigned one district office staff person as SLC project coordinator, one of whose responsibilities was to assist with implementation of the ERO program and study. Based on data contained in the district SLC grants it is estimated that the project coordinator devoted 8.75 percent of a full-time equivalency to each school.

[161]As described in Chapter 1, training activities included: (1) a five-day training course in Washington D.C., in the summer before Year 1 of implementation, which was provided by the program developers and attended by one ERO teacher per school and one program coordinator per district; (2) a shorter training

(continued)

  - Program-specific training, training materials, and ongoing coaching
  - Teacher lodging and meal expenses
  - Meeting costs (meeting room rental, audiovisual equipment, hotel fees, and coordination staff time)

- **Travel:** Costs of travel for ERO teachers, project coordinators, and program developers to attend initial training and midyear training
- **Equipment/supplies:** Training materials, curriculum guides, student materials, classroom libraries, etc.
- **Other direct costs:** Communications, printing, photocopying, and technical support
- **Indirect costs:** Overhead dollars spent on coordinating implementation support and professional development

Disaggregating the cost of the ERO programs by resource makes it possible to identify the factors that have the greatest impact on total cost. This, in turn, makes it possible to understand how total costs would differ were the resource levels and/or unit cost of these resources altered to fit local circumstances in a nonstudy setting.[162]

Table 6.1 presents findings on the cost of the ERO programs per student served (in 2008 dollars).[163] Costs per student are disaggregated by resource and are provided for each implementation year and in total across both years. The last column of Table 6.1 shows how the cost of the ERO programs is distributed across the different resource components. The cost estimates in this table are for both reading programs together (Reading Apprenticeship Academic Literacy and Xtreme Reading).[164]

---

program during the summer before Year 2 of implementation; (3) and two-day "booster" training sessions during each year of the study. Additional training was also provided for replacement teachers in Year 2 of the study. Program coaches visited each ERO teacher approximately five times during each school year.

[162]The cost study uses a Resource Cost Model (RCM) approach to estimate the costs of the programs. Appendix L provides further detail on the data sources and methods used for the cost study.

[163]Average per-student costs across all schools were obtained by summing costs across all schools and dividing by the total number of students enrolled in the study schools (which ranges from 32 to 72). See Appendix L for details on adjustments made to the cost data to account for inflation and other factors.

[164]Program-specific cost estimates are not shown because they would not provide an accurate or "fair" comparison of the two programs' costs. With respect to *salary expenditures* (which represent the largest portion of costs), between-program differences in cost estimates would be due to differences in teacher

(continued)

The Enhanced Reading Opportunities Study

Table 6.1

Cost per Student of the ERO Programs, by Implementation Year (2008-2009 Dollars)

Cost Category	Average Across Both Years ($)	Year 1 ($)	Year 2 ($)	Share of Average Cost per Student (%)
Personnel	1,073	1,085	1,061	55.5
Fringe benefits	321	313	328	16.6
Travel	114	134	95	5.9
Training	260	294	226	13.5
Equipment/supplies	57	80	34	2.9
Other direct costs	37	40	34	1.9
Indirect costs	69	88	51	3.6
Total	1,931	2,035	1,829	

SOURCES: AIR calculations derived from the school-level budget blueprint provided in the Statement of Roles and Responsibilities for Participation in the Enhanced Reading Opportunities Study, salary step schedules downloaded from district Web sites, teacher qualifications from the Enhanced Reading Opportunity teacher survey, the National Center for Education Statistics (NCES) Common Core of Data (CCD) Fiscal Survey (F-33 Data), NCES CWI data, BLS Employment Cost Index for Elementary and Secondary Schools, CPI for all Urban Consumers, and Internal AIR Records of Costs of Training Replacement Teachers.

NOTE: Rounding may cause slight discrepancies in calculating sums and differences.

- The average cost of the one-year programs was $1,931 per student; salary expenditures represent 72 percent of this cost.

Table 6.1 shows that the cost of the ERO programs was larger in the first year of implementation ($2,035 per student) than in the second year ($1,829 per student). This difference is due to the "startup" costs that had to be incurred in the first year, such as initial teacher training and the purchase of materials and equipment (which are reusable and did not need to be

credentials rather than to differences in program costs (to see this, consider that if a school were to implement one of the ERO programs, salary costs would be the same regardless of which program was selected). Presenting cost estimates of *nonsalary expenditures* at the program level would also be of limited utility, because both developers tailored their program budgets to fit the study budget, this amount being the same for both programs. Therefore, by virtue of the study parameters, nonpersonnel expenditures do not differ substantially across schools and therefore programs.

purchased again in the second year). In contrast, recurring costs such as personnel salaries are similar across both years of implementation.

Like most high school courses, salary expenditures represent the largest portion of the cost per student of the ERO programs (72 percent). An additional 13.4 percent of the per-student cost was spent on training resources, while 5.9 percent paid for travel to and from training activities. The remaining 8.5 percent covered equipment/supplies, other direct costs, and indirect costs.[165]

- The cost of the ERO programs was greater in some schools than others, ranging from a high of $2,862 per student to a low of $1,449 per student.

Though the average cost of the programs is $1,931 per student, this average masks differences in the per-student cost of the program across the study schools. The largest part of this variation is attributable to between-school differences in personnel salary costs and fringe benefit expenditures.[166] In contrast, expenditures on nonsalary program components — training, travel, equipment/supplies, and other direct and indirect costs — were similar across the study schools. This suggests that in order to understand the factors that drive the total per-student cost of the ERO programs, it is important to take a closer look at the factors that affect salary expenditures.

To this end, the study team examined the relationship between program costs and the two following factors: (1) the credentials of the ERO teacher (which determines teacher salaries) and (2) program enrollment (since higher program enrollment reduces the amount spent per student, given fixed costs). The findings from this analysis confirm that:

- On average, schools with the highest per-student cost had an ERO teacher with a higher level of educational attainment and more years of teaching experience.
- On average, schools that enrolled fewer students in the ERO program spent more per student, because the total cost of the programs is spread over a smaller number of students.[167]

---

[165]See Appendix L for a more detailed breakdown of nonsalary program costs (Table M.1).

[166]Personnel salary expenditures ranged from $1,579 per student for the highest-spending schools compared with $809 for the lowest-spending school; fringe benefits ranged from $690 for the highest-spending school compared with $176 for the lowest-spending school.

[167]See Appendix L for more detailed findings from this analysis.

This means that the cost per student of implementing the ERO programs depends primarily on the educational attainment and years of service of the teacher(s) assigned to the reading class and the number of students enrolled.

These findings are used in the final section of the chapter to simulate the potential cost of implementing the ERO programs in a nonstudy setting. However, a key factor in these simulations is how the ERO programs might be implemented without the constraints and monitoring that came from participating in a research demonstration. To this end, the following section examines the ways in which the ERO high schools modified their implementation of the programs after the end of the study.

Poststudy Continuation and Modification of the ERO Programs in the Study Schools

During the spring and early summer of 2009, the study team interviewed staff in the study schools to determine the status of adolescent literacy programming in their school after the conclusion of the grant-required ERO implementation period.[168] The study team was able to interview school-level staff in 30 of the 34 study schools,[169] all of whom were familiar with the ERO study as it was implemented in their schools.[170]

There were two primary questions that guided these interviews. The first question was whether schools in the ERO study continued to use their ERO program in the school years after the study ended (see Figure 6.1). The answer to this question provides information about whether participating schools found enough value in the programs that they continued to support them on their own.[171] The second key question was how, if the districts and schools sustained them, the programs were modified or adapted in the poststudy period (see Figure 6.2).

---

[168]As discussed in Chapter 1, schools received funding for their Smaller Learning Communities through a grant from the U.S. Department of Education's OVAE. As part of this grant, they were required to implement RAAL or Xtreme Reading in 2005-2006 and 2006-2007.

[169]The study team was unable to schedule interviews in four schools because summer breaks had started in those districts at the time schools were contacted.

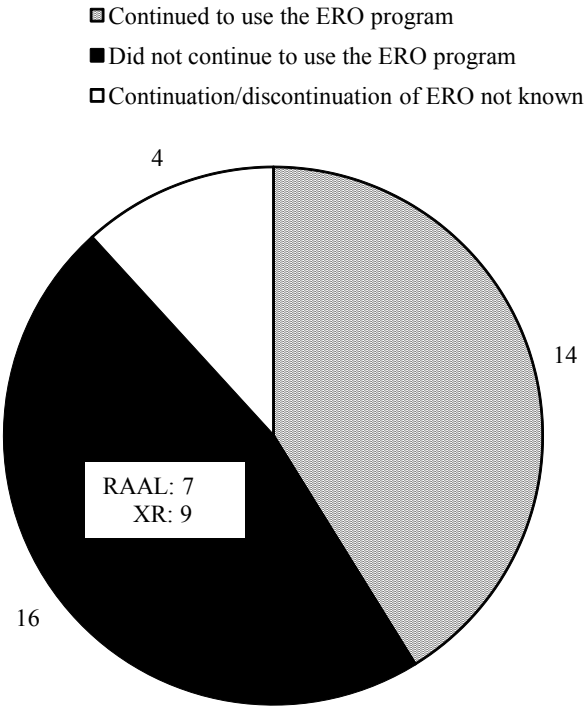
[170]The study team also conducted interviews with district-level staff to obtain a broader perspective on adolescent literacy programming in the school district. Findings from these interviews are presented in Appendix M. See also Appendix M for a discussion of the methodology used to collect the interview data (samples, protocols, and analysis).

[171]See Appendix M for findings related to two other aspects of adolescent literacy programming in the study schools during the poststudy period: literacy-focused professional development and funding for reading interventions.

The Enhanced Reading Opportunities Study

Figure 6.1

Continuation of the ERO Programs in the Study Schools After the Study-Required Implementation Period



SOURCES: Interviews conducted with representatives of ERO schools in spring and summer 2009. Program-specific counts are not disclosed in two of the categories, to protect confidentiality.

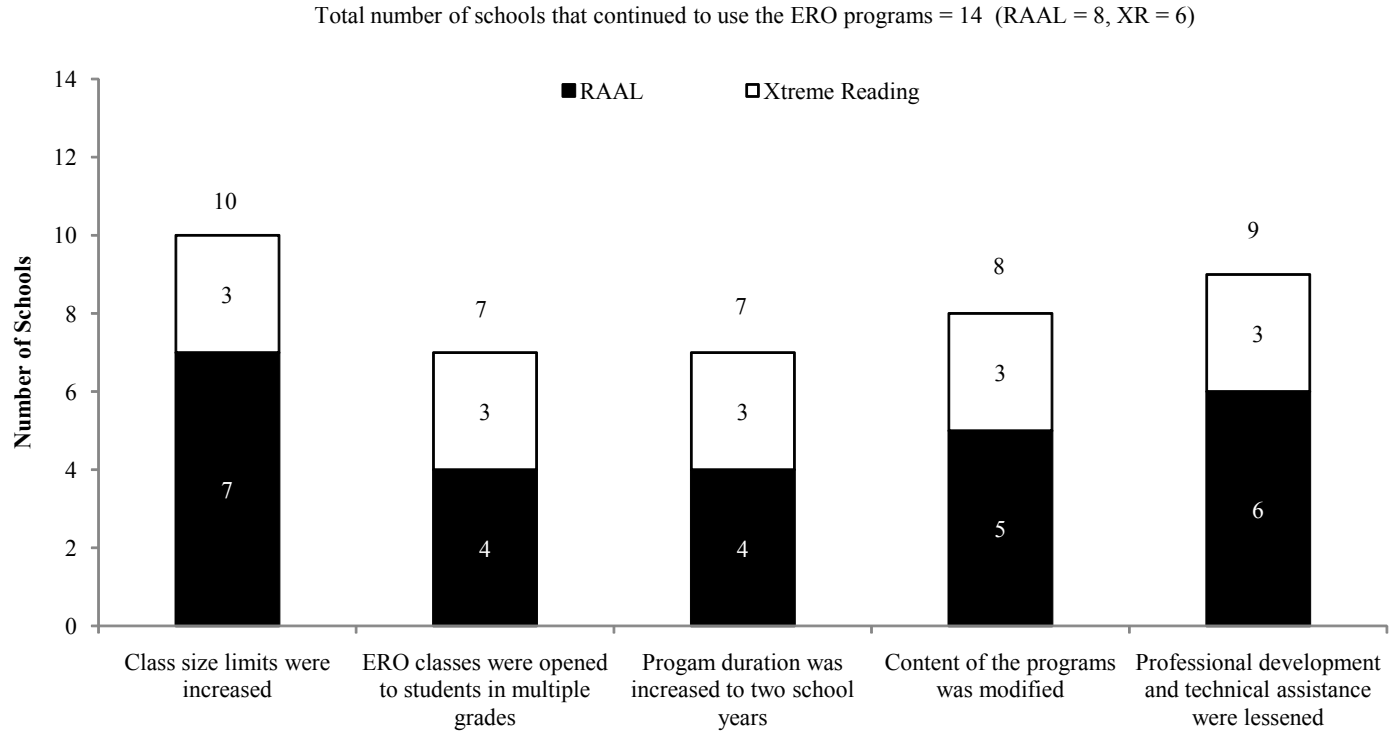
- Fourteen high schools (47 percent of interviewed schools) continued to offer the ERO program after the end of the study-required implementation period.[172]

[172]At the district level, the ERO programs were still in use in five of 10 study districts as of the end of the 2008-2009 school year. The decision to continue/discontinue the ERO program was most often made by districts (four districts) or by individual schools separately (four districts); in the remaining two districts, the decision was made by the district and school together. Appendix M provides further information on the continuation/discontinuation decisions of each school district.

The Enhanced Reading Opportunities Study

Figure 6.2

Most Common Modifications to the ERO Programs in Schools That Continued to Use the Programs After the Study-Required Implementation Period



SOURCES: Interviews conducted with representatives of ERO schools in spring and summer 2009.

- Most of the modifications to the ERO programs expanded the reach of the programs to serve more students or provide longer-term support to students.

For the purposes of the ERO study, implementation parameters were set on the target population and the number of students served by the literacy programs. English Language Learning (ELL) and special education students who required specific classroom, instructional, or testing accommodations were not eligible for the ERO classes.[173] The programs served ninth-grade students only and were offered to student for only one school year. In addition, class sizes were capped at a maximum of 15 students, and only one teacher in each school was trained to teach the ERO class.

After the study ended, however, interviewees in the 14 high schools that continued to offer the ERO programs reported that they had deviated from these study-based parameters. The most common modification was to increase the ERO class size beyond the 15-student study cap (10 schools, or 71 percent of continuing schools). In seven schools (50 percent of the continuing schools), the ERO classes were opened up to students in multiple grades. Seven schools (50 percent of continuing schools) reported that they extended the duration of the ERO program from one to two academic years, such that students continued to receive instruction in ERO strategies in tenth grade as well.[174] The primary reason cited for this change was to provide added reinforcement for students through additional support and ongoing attention to the reading strategies the programs emphasize. Other less common modifications made by schools to increase program capacity were to expand the eligibility criteria to include ELL and special education students and to have another ERO teacher trained by the program developers.

- In eight schools (57 percent of continuing schools), modifications were made to the content of the ERO programs after the study ended.

During the two study-required implementation years, teachers had agreed to follow the programs' detailed lesson plans and curriculum guides, to use equipment specified by the developers, and to teach using program-specific resources.[175] After the study ended, however, eight of the 14 schools that continued to use the ERO programs reported that they had made

---

[173]The ERO programs were not designed to accommodate the special needs of these students or the potential scheduling conflicts with other services that the students were likely to receive.

[174]For example, some schools adjusted program duration by extending the content to a multiyear ERO course. Another modification was to offer the course with "looping," an approach in which cohorts of students who received the program in ninth grade stayed together for regular English language arts in grades 10 and 11, receiving reinforcement in the ERO strategies from their ELA teachers.

[175]The two programs differed in the level of specificity of their instructional materials, but both provided goals for what teachers were supposed to teach and the time frame for instructional units.

modifications to the instructional models. In particular, five schools had added new content to the course and six schools had eliminated portions of the original curriculum. The content added to the programs included adding instruction in test preparation strategies or grammar or supplying new books to augment the developers' original classroom libraries. Content that was eliminated included student practice work or scripted sections of the lessons.

- In nine schools (64 percent of continuing schools), levels of professional development and technical assistance were reduced.

During the ERO study implementation years, professional development through summer institutes, midyear "booster" sessions, and on-site coaching had been prominent features of the program. However, interviewees at nine schools reported that their schools provided no additional program-specific training after the end of the study-required implementation period.[176]

These modifications may affect program impacts, and some certainly influence costs. The potential effect of these modifications on program impacts cannot be estimated as part of this study.[177] However, the cost implications of these modifications are explored in the next section, which presents estimates of the cost of the programs in a hypothetical nonstudy setting.

Estimated Program Costs Under Nonstudy Conditions

The ERO study evaluated two program models as they were intended to be delivered, and therefore set parameters on the implementation of the programs in the study schools (amount and location of the trainings, program materials, class size, and teachers' schedules).[178] It is likely that in a nonstudy setting, costs could be different because school administrators would set different parameters on the implementation of the reading interventions. In fact, as reported in the prior section of the chapter, schools did modify their implementation of the ERO programs.

---

[176]However, the program developers invited former ERO teachers from three of the schools to become program consultants or staff developers and provided them with the additional training needed for these new roles.

[177]This evaluation's reported impact findings on student outcomes are for programs implemented within the parameters set by the study. Variation from these parameters — that is, modifications like those made by schools that continued the ERO programs — could result in a different pattern of impacts.

[178]For example, class sizes were capped at 15 students, and ERO teachers were assigned to teach four sessions of the program rather than carrying a five-course load. (Class schedule data from the study schools indicates that under nonstudy conditions, a teacher would have to be responsible for teaching five class sections to be considered full time).

In order to illustrate how the cost of the ERO program could differ in a nonstudy setting, the study team calculated cost figures for a hypothetical school in the study districts that chooses to implement the reading intervention. These nonstudy cost estimates are based on assumptions guided by the previous sections' findings about how the study schools chose to implement the programs after the end of the required period:

- **Student enrollment:** As reported in the previous section, most of the modifications reported by schools that continued to offer the ERO programs focused on increasing the number of students served. Because the number of students served is an important determinant of the per-student cost of the programs, cost estimates are presented for three levels of student enrollment: 60 students, 90 students, and 120 students. An enrollment of 60 students represents the original study size (15 students in four class sections); an enrollment of 90 students represents the approximate number of students estimated to be eligible for the programs in the study schools (that is, the average number of students in the ERO and the non-ERO group in study schools),[179] while an enrollment of 120 students reflects the finding that some schools increased their class size to 30 students after the end of the study (which would amount to 120 students across four sections).

- **Personnel:** It is assumed that the ERO classes are taught by one full-time ERO teacher (since 12 of the 14 schools that continued to implement the ERO programs did not train an additional teacher).[180] As discussed in the first section of this chapter, the qualifications of the ERO teacher are the most important driver of the cost of the programs; thus, the credentials and experience of this teacher are allowed to vary in the cost simulations.[181,182]

[179]In the first year of the study, an average of 86 students per school were eligible for the ERO programs (that is, these students were two to five years below grade level), and in the second year of the study, 79 students per school were eligible.

[180]It is also assumed that the same teacher would teach all sections of the ERO class.

[181]Salaries used for these cost calculations are based on the average salary across the study districts for instructors with a given number of years of experience and education. Fringe benefits are based on the average fringe rate across the study districts. See Appendix L for details on assumptions about the salaries associated with different levels of experience and education (Appendix Table L.3).

[182]The cost simulations also assume that there is no district-based ERO project coordinator, because in a nonstudy setting, coordination of the ERO program could be subsumed under the responsibilities of the director of literacy programming at the district or school.

- **Training, travel, and equipment/supplies:** These costs are assumed to be at the same level as in the ERO study, but they are spread out over a four-year period rather than a two-year period.[183] This assumption is based on the finding that after the end of the study, nine of the 14 high schools that continued the ERO programs did not provide additional professional development support. In addition, given that equipment and supplies are less than 3 percent of program costs, it was assumed that the costs of replenishing supplies or maintaining equipment would be negligible and absorbed into the overall school budget.

Table 6.2 presents cost estimates based on these assumptions about the nonstudy setting. In order to interpret the findings in this table, it is useful to start with the conditions that are closest to the "study setting," that is, that the ERO teacher has 10 years experience and a master's degree (which represents the "average" ERO teacher in the study[184]) and that program enrollment is 60 students. In this scenario, the per-student cost of implementing the programs would be $1,391, which is 72 percent of the per-student cost of implementing these programs in the study setting ($1,931 per student). This reduction in cost is due to the fact that in a nonstudy setting, upfront program costs (such as equipment and supplies) can be smoothed out over a longer period.

As also seen in Table 6.2, the estimated cost of the programs varies depending on the credentials of the ERO teacher and total program enrollment. For example, if the ERO teacher's credentials were as above (a master's degree and 10 years experience), but program enrollment were increased to 90 students, then the per-student cost of implementing the programs would be $927, which is 48 percent of the per-student cost of implementing these programs in the study setting ($1,931 per student). More broadly, the projected cost estimates range from $1,801 per student at a school with low enrollment served by a highly credentialed

[183]This assumes that the ERO teacher remains in the schools for four full years and that schools continue to implement the same ERO program for four full years.

[184]Teachers across both study years had an average of 11.4 years of experience, while the majority had obtained a master's degree (55.9 percent had a master's degree, 27.9 percent had a bachelor's degree, and 16.2 percent had a doctorate).

The Enhanced Reading Opportunities Study

Table 6.2

Projected Cost per Student of the ERO Programs in a Hypothetical Nonstudy Setting, by Teacher Qualifications and Total Student Enrollment (2008-2009 Dollars)

Total Students	Highest Degree Obtained	0-1 Year of Experience ($)	10 Years of Experience ($)	20 Years of Experience ($)
	Bachelor's	570	657	744
120	Master's	603	695	806
	Doctorate	650	767	900
	Bachelor's	761	876	992
90	Master's	804	927	1,075
	Doctorate	867	1,023	1,201
	Bachelor's	1,141	1,314	1,488
60	Master's	1,206	1,391	1,613
	Doctorate	1,300	1,535	1,801

SOURCES: AIR calculations derived from the school-level budget blueprint provided in the Statement of Roles and Responsibilities for Participation in the Enhanced Reading Opportunities Study, salary step schedules downloaded from district Web sites, the National Center for Education Statistics (NCES) Common Core of Data (CCD) Fiscal Survey (F-33 Data), NCES CWI data, BLS Employment Cost Index for Elementary and Secondary Schools, and the CPI for all Urban Consumers.

NOTES: Table describes spending on the average student participating in the Enhanced Reading Opportunities (ERO) program in a hypothetical nonstudy setting, assuming one full-time ERO instructor, 60 students enrolled, and costs annualized over four years.

Rounding may cause slight discrepancies in calculating sums and differences.

teacher, to $570 per student for a school with high enrollment and a less-credentialed teacher. These projected costs are comparable with recent estimates of the cost of a high school remedial reading class; for example, the average per-student cost of remedial reading courses in Ohio is $1,321 in 2008 dollars.[185]

[185]Chambers (1999) showed the average per-pupil cost of remedial reading courses across 442 Ohio high schools to be $860 in 1995, which is equivalent to $1,321 in 2008 dollars.

References

Allensworth, Elaine M., and John Q. Easton. 2007. *What Matters for Staying On-Track and Graduating in Chicago Public High Schools: A Close Look at Course Grades, Failure, and Attendance in the Freshman Year.* Chicago: Consortium on Chicago School Research at the University of Chicago.

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 upon request at ERO@air.org.

Balfanz, Robert, and Nettie Legters 2004. *Locating the Dropout Crisis.* Baltimore, MD: Center for Social Organization of Schools. Web site: http://www.csos.jhu.edu/tdhs/rsch/Locating_Dropouts.pdf.

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, Howard. 2006. *The Core Analytics of Randomized Experiments for Social Research*. Working Paper. New York: MDRC.

Carnevale, Anthony P. 2001. *Help Wanted…College Required.* Washington, DC: Educational Testing Service, Office for Public Leadership.

Chambers, Jay G. 1999. *Measuring Resources in Education: From Accounting to the Resource Cost Model Approach.* Washington, DC: U.S. Department of Education, National Center for Education Statistics, Working Paper Series. Web site: http://nces.ed.gov/pubs99/199916.pdf).

Corrin, William, Marie-Andree Somers, James J. Kemple, J., Elizabeth Nelson, and Susan Sepanik. 2008. *The Enhanced Reading Opportunities Study: Findings from the Second Year of Implementation* (NCEE 2009-4036). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Diamond, John B, William J. Corrin, and Judith Levinson. 2004. *Challenging the Achievement Gap in a Suburban High School: A Multimethod Analysis of an Adolescent Literacy Initiative.* Naperville, IL: Learning Point Associates. Web site: http://www.ncrel.org/gap/studies/msanreport.htm.

Diana Oxley (ed.). 2006. *Critical Issues in Development and Implementation of High School Small Learning Communities.* Portland, OR: Northwest Educational Laboratory.

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.

Hill, Carolyn J., Howard S. Bloom, Alison R. Black, and Mark W. Lipsey. 2008. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives* 2, 3: 172-177.

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

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2006. *Experimental Analysis of Neighborhood Effects.* (Web appendix). Web site: http://www.nber.org/mtopublic/mto_exp_a.pdf.

Levin, Henry M., Doran Catlin, and Alex Elson. 2007. "Costs of Implementing Adolescent Literacy Programs." Pages 61-91 in Donald D. Deshler, Annemarie Sullivan Palincsar, Gina Biancarosa, and Marnie Nair, *Informed Choices for Struggling Adolescent Readers: A Research-Based Guide to Instructional Programs and Practices*. Newark, DE: International Reading Association.

Levin, Henry M, Doran Catlin, and Alex Elson. 2009. *Adolescent Literacy Programs: Costs of Implementation.* New York: Carnegie Corporation of New York.

Levin, Henry M., and Patrick J. McEwan. 2001. *Cost Effectiveness Analysis: Methods and Applications.* Thousand Oaks, CA: Sage Publications.

Lipsey, Mark. 1990. *Design Sensitivity: Statistical Power for Experimental Research*. Newbury Park, CA: Sage.

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.

May, Henry, Irma Perez-Johnson, Joshua Haimson, Samina Sattar, and Phil Gleason. 2009. *Using State Tests in Education Experiments: A Discussion of the Issues (Revised Draft Report).* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

NCAA Freshman-Eligibility Standards Quick Reference Sheet. 2008. NCAA Eligibility Center. Web site: https:web1.ncaa.org/eligibilitycenter/hs/d1_standards.pdf.

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.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd Ed.). Sage Publications: Thousand Oaks, CA.

Roderick, Melissa. 1993. The Path to Dropping Out: Evidence for Intervention. Westport, CT: Auburn House.

Schochet, Peter Z. 2008. *Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations* (NCEE 2008-4018). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Schochet, Peter Z. 2008. *Guidelines for Multiple Testing in Experimental Evaluations of Educational Interventions. Final report for the Institute of Education Sciences.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

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 Disabilitie: Strategies for Accessing the Curriculum.* New York: Corwin.

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.

Taylor, Lori L., and William J. Fowler Jr. 2007. *Documentation for the NCES Comparable Wage Index Data Files, 2005* (NCES Report No. 2007397). Web site: https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007397.

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.

Westfall. Peter H., and S. Stanley Young. 1993. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment.* New York: Wiley and Sons.

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.

Appendix A

The ERO Programs and the ERO Teachers

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. Because the intent of the study was not simply to evaluate a specific literacy program, but rather a type of literacy program, the process was designed to select two programs to test for effectiveness. While each of the two programs could be tested individually, together they could be tested as representative of a class of intervention. 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 new programs) and 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 Enhanced Reading Opportunities (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 (RAAL) and the University of Kansas Center for Research on Learning's (KU-CRL) Xtreme Reading.

Overall Goals and Approach

The overarching goals of both RAAL and Xtreme Reading are 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 conducive 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

[1] American Institutes for Research (2004).
[2] Biancarosa and Snow (2004).

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 them 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 RAAL and Xtreme Reading are discussed categorically below. This discussion is based on information provided by the two program developers. Appendix Table A.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]

Developers' Implementation Philosophy

In implementing RAAL, 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 that they expect will motivate their classes. Teachers have flexibility in how they include various aspects of the RAAL 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 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 developers. Xtreme Reading teachers follow a prescribed implementation plan, following specific day-

[3]Additional 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 C provides more information specific to each program.

[4]The 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

Appendix Table A.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	<u>Learning Environment</u> Establish "social reading community" early in program <u>Comprehension Instruction</u> Five curricular strands of classroom instruction: 1. Metacognitive Conversation 2. Silent Sustained Reading 3. Language Study 4. Content/Theme 5. Writing	<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 <u>Comprehension Instruction</u> Focus of rest of course on developing literacy skills through seven learning strategies: 1. LINCS 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: 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: 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

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 RAAL 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 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 more student-centered, as students assume more responsibility for monitoring their own comprehension and adjust their use of skills as needed. In RAAL — 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. Appendix Table A.1 provides an overview of the key elements of each ERO program. Both developers' curriculum designs highlight the equal importance of creating a classroom learning environment that is conducive to learning and focusing instruction on strategies that promote reading comprehension skills and proficiency.

The core of the RAAL 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 improvement.

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 RAAL 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 RAAL.
- **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 implementation 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: LINCS 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 who are 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 Classes and Student Schedules

As discussed in Chapter 1, both RAAL and Xtreme Reading 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. Appendix Table A.2 presents two examples of how the supplemental ERO classes fit into student schedules. 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 80 to 90 minutes every other day (Example 2).[5] Second, since the modal number of course slots in the schools' schedule models was eight slots and the mean was 7.7, Example 1 reflects a schedule with seven course slots, and Example 2 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. Both examples show ERO and non-ERO group students scheduled the same amount of required courses (four in Example 1 and five in Example 2), and the ERO class replaced one of the students' elective courses.

[5]Although most schools employed these two schedule models, there were three schools in the second year of implementation that offered a 4x4 block schedule, where students typically take the same four 90-minute courses every day for an entire term. These students then take four different courses during the second term. One of these schools using a 4x4 block schedule implemented the Xtreme Reading class as a one-semester class that met for 90 minutes every day, with some of the ERO students taking the class the first semester and the rest taking it the second semester. The other two schools, from another district, paired the RAAL class with a year-long computer lab during which the ERO students used TeenBiz, the online application that is part of the RAAL program. The computer lab was not taught by the ERO teacher, and no students from the non-ERO group were enrolled. The ERO class and the computer lab were alternated every other day throughout the year similar to the A/B block schedule model, remaining aligned with the expectations of the study and the grant that the ERO classes be full-year classes. The study team conducted sensitivity testing of the overall impacts and determined that they are not sensitive to the inclusion of these schools, either collectively or independently.

The Enhanced Reading Opportunities Study

Appendix Table A.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	***ERO***	***Elective***
6	Elective	Elective
7	Elective	Elective

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

	ERO Students		Non-ERO Students	
Period	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	***ERO***	Required course	***Elective***
4	Elective	Elective	Elective	Elective

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

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 willing to teach the ERO course 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 study.

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 might have a more realistic chance of identifying 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.

Appendix Table A.3a and A.3b provide a list of background characteristics for the teachers in each of the two ERO programs in each year of implementation.[6] The average number of years of experience for ERO teachers was between 11 and 12 years (11.2 years in the first year and 11.9 years in the second year), although teaching experience ranged from student teaching to over 30 years as a regular classroom teacher. Over two-thirds of the teachers had graduate-level degrees (73.5 percent and 67.7 percent in the first and second years, respectively), and almost all held high school-level certification (97.1 percent in the first year, and 100 percent in the second year). The majority of the teachers were certified in English language arts (76.5 percent and 85.3 percent in the first and second years respectively). Nearly 18 percent in the first year and nearly 9 percent in the second year held social studies certification and 6 percent held certification in some other area (both years). Teachers reported attending an average of more than 40 hours of professional development in the two years prior to the beginning of the ERO program (45.4 hours in the first year and 41.4 hours in the second year).[7]

The differences in teacher characteristics between the first and second years reflect the fact that not all the schools retained the same ERO teacher across both years of implementation. Of the 34 ERO teachers in the second year of program implementation, 25 had returned after having taught all of the first year of the program. Two of the teachers replaced first-year teachers in the middle of the first year. Seven teachers replaced first-year teachers who left at

[6]Information in these tables 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.

[7]Differences 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. The impact analyses conducted for this report and the prior reports account for differences across school characteristics (and, thus, across teachers) by including regression covariates for each school.

The Enhanced Reading Opportunities Study

Appendix Table A.3a

Background Characteristics of ERO Teachers, First Year of Implementation

Characteristic	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
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
Certified in English language arts subject matter (%)	76.5	70.6	82.4
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 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.

[a]Missing data: One to two teachers did not respond.

[b]Missing data: Four to five teachers did not respond.

The Enhanced Reading Opportunities Study

Appendix Table A.3b

Background Characteristics of ERO Teachers, Second Year of Implementation

Characteristic	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Gender (%)			
Male	11.8	0.0	23.5
Female	88.2	100.0	76.5
Total time teaching (years)	11.9	11.3	12.4
Total time teaching at current school (years)	5.4	5.2	5.6
Total time teaching at current level (years)[a]	8.0	8.0	8.0
Total time teaching English language arts or social studies (years)[a]	10.3	10.1	10.6
Master's degree or higher at the start of ERO program (%)	67.7	64.7	70.6
Holds high school-level teaching certification (%)	100.0	100.0	100.0
Certified in English language arts subject matter (%)	85.3	76.5	94.1
Number of professional development workshops attended in the last two years prior to becoming an ERO teacher[a]	4.2	4.7	3.7
Number of hours spent in professional development workshops during the last two years prior to becoming an ERO teacher[b]	41.4	35.5	48.2
Taught the ERO class for the full school year (%)	100.0	100.0	100.0
Taught the ERO class in its first year (%)	79.4	76.5	82.4
Sample size	34	17	17

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

NOTES: This table contains data from the baseline teacher survey administered to teachers at their point of entry into the ERO demonstration.

Rounding may cause slight discrepancies in calculating sums and differences.

[a]Missing data: One teacher did not respond.

[b]Missing data: Six teachers did not respond.

the end of the first year.[8] Given that the replacement teachers came to the second year of implementation with no experience with the ERO programs, there were differences in their training and support compared with that of the returning teachers. These differences are discussed in the next section of the appendix. Also, differences in implementation findings for the nine replacement teachers, the teachers they replaced, and the teachers who participated in both years of the study were discussed in Chapter 3.

Training and Technical Assistance

Training and technical assistance were delivered to the ERO teachers in several ways. ERO teachers attended summer training institutes held prior to each year of implementation (five days the first summer and three days the second summer). In addition, RAAL teachers attended two two-day booster training sessions, and Xtreme Reading teachers attended one two-day booster training each year. The program developers also provided onsite support to the ERO teachers through instructional coaching visits to the schools both years. RAAL teachers received ongoing support through access to a special online listserv that was set up by the developer for the project. District program coordinators were invited to observe (and some did attend) the trainings to familiarize them with the programs in case they had to provide technical assistance or other support to ERO teachers.

Additional training was provided to teachers new to the ERO programs in the second year. Prior to the second summer training, new RAAL teachers participated in a national Reading Apprenticeship Institute to learn the central strategies and philosophies of the Reading Apprenticeship program.[9] New RAAL teachers were also offered up to five extra days of coaching support if the teacher, coach, and the district coordinator determined that supplemental technical assistance was needed. New Xtreme Reading teachers attended a two-day training immediately prior to the three-day training for all Xtreme Reading teachers (new and returnees from the first year) during the summer before school started in the 2006-2007 school year. Technical assistance for both programs is described in more detail below. In addition, Appen-

[8]Replacement teachers were identified by the schools and the districts. Their résumés were submitted to the study team for initial screening. The study team screened them for a teaching certification and at least two years of teaching experience. Résumés that satisfied those criteria were forwarded to ED for final approval.

[9]The Reading Apprenticeship Academic Literacy (RAAL) 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, new ERO teachers received additional training that addressed aspects of Reading Apprenticeship that are specific to the ERO study. A specific supplemental training that focused entirely on RAAL was not offered to new teachers because there were very few teachers that were replaced at the end of Year 1, and the economies of scale found in the first-year start up conditions could not be replicated.

dix Table A.4 summarizes the activities each of the developers provided to the ERO teachers for the 2005-2006 and 2006-2007 school years. Figure 1.2 in Chapter 1 provided a timeline for both years of program implementation, including teacher selection, teacher replacement, and teacher training.

Summer Trainings

The summer teacher training institutes for both programs were conducted in August 2005 in the first year and in late July/early August 2006 in the second year. The RAAL training was conducted by the program developer, experienced RAAL teachers (teachers who are employees of WestEd),[10] and the coaches who would work with the ERO teachers throughout each school year. In the first summer, the training provided an introduction to the RAAL course as a whole and also focused on curricular units to be taught during the first portion of the year. The second summer training focused on those program areas that were identified by returning ERO teachers or the developers as needing additional support and instruction. The staff at WestEd determined areas of teacher weakness during their coaching visits to the RAAL classrooms throughout the first year.

The RAAL trainings included modeling and whole-group discussions led by the developers, the coaches, and selected returning teachers. There were also small group activities such as roundtable discussions on particular subjects and individual planning periods. Teachers were able to meet with the RAAL coaches individually during the training, as well. In the first year, 15 of the 17 RAAL teachers attended the summer training, and the other two attended national Reading Apprenticeship workshops before they started teaching the course.[11] In the second year, fourteen of the 17 RAAL teachers attended the summer training.

The Xtreme Reading summer training institutes were conducted by the program developers, research staff from the University of Kansas Center for Research on Learning, a behavioral specialist (in the second summer), and the coaches who would work with the teachers. In the first summer, the training provided an introduction to the Xtreme Reading program and also addressed the curricular units to be taught in the beginning of the year. The second summer

[10]Some of the returning teachers from the study schools who taught all of the first year were asked to lead presentations on particular units during the second summer institute.

[11]The Reading Apprenticeship Academic Literacy course implemented as part of 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 specific to the ERO study.

The Enhanced Reading Opportunities Study

Appendix Table A.4

Training and Technical Assistance Provided During Each School Year, by ERO Program

2005-2006 School Year

	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

2006-2007 School Year

	Summer Training	School-Year Booster Training	Additional Supports
Reading Apprenticeship	One 3-day training for new and returning teachers (August)	Two 2-day trainings (October; March)	Three 2-day on-site coaching visits Weekly e-mail and phone calls Listserv
Xtreme Reading	One 2-day training for new teachers, followed by a 3-day training for both new and returning teachers (July-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

training mainly addressed challenges faced by ERO teachers during the first year of implementation, though the teaching of new strategies was also a part of the agenda. KU-CRL staff worked with ERO teachers during the Year 1 booster training to identify the challenges that would be discussed at the summer training. Training methods 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. All of the 17 Xtreme Reading teachers attended both summer training institutes.

Booster Trainings

The school-year booster trainings during each implementation year (two for RAAL and one for Xtreme Reading, each year) were conducted in a similar format to the summer training institutes and were two days each in duration. The program developers used these trainings to address challenges the teachers were facing in the classes, refine teaching approaches, and introduce new material and curriculum changes. 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 regarding the implementation of the program that had come up during the first part of the year. In the first year, all 17 RAAL teachers attended both training sessions. Sixteen of the 17 Xtreme Reading teachers attended the midyear booster session in person, and one teacher participated by telephone. In the second year, 15 of the 17 RAAL teachers attended the fall booster training session and 14 attended the spring session. Fifteen of 17 Xtreme Reading teachers attended the booster training session in person, and two teachers participated by telephone.

Ongoing Technical Assistance

Both programs provided on-site coaching and electronic and telephone communication among teachers and their coaches. RAAL also made a listserv available to teachers. Each year the RAAL and Xtreme Reading coaches made three two-day visits to each of the teachers. During these visits 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 cases of teacher turnover during the first year of implementation, coaches provided additional technical assistance to the replacement teachers.

Appendix B

ERO Student Survey Measures

As described in Chapter 2, a student survey was administered to each cohort of ninth-grade study participants near the end of their ninth-grade year (spring 2006 for Cohort 1 and spring 2007 for Cohort 2). The questions in the 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 reading activities.

This appendix describes the measures constructed from students' responses to the survey. In general, 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 survey is included at the end of the appendix.

Measures of Self-Reported Participation in Supplemental Literacy Support Activities

This section 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 of school, and (4) working with a reading or writing tutor outside of 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 5, 8, 11, and 14) 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 6, 9, 12, and 15) 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 7, 10, 13, and 16) asks about the frequency of this support. The response choices for this item 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"

Combining responses to these three items, a measure was constructed 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 5, 8, 11 or 14, the participation measure for the activity was coded to zero (0). For students who answered "yes" to questions 5, 8, 11 or 14, Appendix Table B.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 6, 9, 12, and 15). The rows represent frequency — "how often" a student received that help (questions 7, 10, 13, 16). 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 student survey included 18 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 likely to include extended passages. The team also focused on groups of items for which student responses were highly correlated (that is, groups of items that were correlated with Cronbach's alpha > .70). The seven items used to construct a measure of in-school reading frequency were correlated with Cronbach's alpha = .83 for Cohort 1 and Cronbach's alpha = .71 for Cohort 2, and the seven items used to construct a measure of out-of-school reading were correlated with Cronbach's alpha = .73 for Cohort 1 and Cronbach's alpha = .75 for Cohort 2.

The Enhanced Reading Opportunities Study

Appendix Table B.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 (x 0.1)	0.4	0.8	1.8	2.7	3.6
Once a month (x 0.25)	1	2	4.5	6.75	9
Every other week (x 0.5)	2	4	9	13.5	18
Once a week (x 1)	4	8	18	27	36
Twice a week (x 2)	8	16	36	54	72
Three to four times a week (x 3.5)	14	28	63	94.5	126
Every day (x 5)	20	40	90	135	180

The study team also developed a measure of the frequency with which students used reading strategies in their reading for other courses. The six strategies included in the measure are often used by proficient readers and are incorporated into the instruction of the two supplemental literacy programs for this study.[1]

Frequency of In-School Reading

7 items, Cronbach's Alpha = .83 for Cohort 1, .71 for Cohort 2

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 is

[1]Biancarosa and Snow, 2004.

imputed using the mean of the values for the other items. If more than three of the items were missing, the entire construct is coded as missing for a given student.

Question 3. Please indicate about how OFTEN, during the past month, you READ each of the following in class or for homework.

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
f. Newspaper or magazine articles
g. 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 for Cohort 1, .75 for Cohort 2

This construct is designed to measure the frequency with which students read extended texts outside of 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 values for the other items. If more than four of the items were missing, the entire construct was coded as missing.

Question 4. During the <u>past month</u>, about how OFTEN did you READ each of the following when you were not in school and not doing homework?

a. Fiction books or stories
b. Poetry
d. Biographies or autobiographies
e. Books about science

f. Books about history
g. Newspaper or magazine articles
h. 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

4 items, Cronbach's Alpha = .88 for Cohort 1, .77 for Cohort 2

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 are both consistent with the strategies taught by the ERO programs and seen as antecedents to reading proficiency. The two questions that comprise this measure were asked in the context of the reading that students do for their English class and the reading they do for one other core content area class (history, science, or math) for a total of 4 items.[2]

Question 17. 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.

b. When I'm reading for English class I stop once in a while and go over what I have read.

Scale: 1 = "Strongly Disagree" to 4 = "Strongly Agree"

[2]The items used to calculate this construct changed from the first cohort to the second cohort. The construct reported on in the second report (Corrin et al., 2008) is included here. See Kemple et al. (2008) for the alternative calculation of this construct. Thus, please note that the Cronbach's alphas reported for this construct for each cohort represent different item configurations.

Question 18. For which one of the following classes did you do the most reading during the past school year?

1. History (or Social Studies)
2. Science
3. Math

Question 19. Please indicate how much you DISAGREE or AGREE with the following statements about the class you chose in Question 18.

a. I ask myself questions to make sure I know the material that I have been studying for class.

b. When I'm reading for class I stop once in a while and go over what I have read.

Scale: 1 = "Strongly Disagree" to 4 = "Strongly Agree"

ENHANCED READING OPPORTUNITIES

STUDENT FOLLOW-UP QUESTIONNAIRE
SPRING 2007
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**

The first question asks you about your future education.

(1) 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.

Please mark (✓) one answer on each line.

(2) 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.	1❍	2❍	3❍	4❍
b.	Reading is one of my favorite activities.	1❍	2❍	3❍	4❍
c.	Writing things like stories or letters is one of my favorite activities.	1❍	2❍	3❍	4❍
d.	Writing helps me share my ideas.	1❍	2❍	3❍	4❍
e.	When I have free time, I rarely choose to read over doing other activities.	1❍	2❍	3❍	4❍
f.	I read because it helps me do better in my classes.	1❍	2❍	3❍	4❍
g.	I read to see what is going on in the world, the country, and/or my community.	1❍	2❍	3❍	4❍
h.	I read because I enjoy it.	1❍	2❍	3❍	4❍
i.	I read in order to learn new things.	1❍	2❍	3❍	4❍
j.	I read to learn how other people see things.	1❍	2❍	3❍	4❍

Continue ➡

The next question asks about what you read for school.

(3) Please indicate about how OFTEN, during the past month, you READ each of the following in class or for homework. 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	2	3	4	5	6	7
b.	Science textbook	1	2	3	4	5	6	7
c.	Math textbook	1	2	3	4	5	6	7
d.	Novels, short stories, plays, poetry, or essays	1	2	3	4	5	6	7
e.	Research papers, reports, graphs, charts, or tables	1	2	3	4	5	6	7
f.	Newspaper or magazine articles	1	2	3	4	5	6	7
g.	Websites on the Internet	1	2	3	4	5	6	7
h.	Workbooks	1	2	3	4	5	6	7

This section is about reading you do that is not for school.

Please mark (✓) one answer on each line.

(4) During the past month, about how OFTEN did you READ each of the following, when you were not in school and not doing homework?

		Never	At least once	Every other week	Once a week	Twice a week	3-4 times a week	Every day
a.	Fiction books or stories (books or stories about imagined events)	1	2	3	4	5	6	7
b.	Plays	1	2	3	4	5	6	7
c.	Poetry	1	2	3	4	5	6	7
d.	Biographies or autobiographies	1	2	3	4	5	6	7
e.	Books about science (for example, nature, animals, astronomy)	1	2	3	4	5	6	7
f.	Books about history	1	2	3	4	5	6	7
g.	Newspaper or magazine articles	1	2	3	4	5	6	7
h.	Religious books (e.g., Koran, Bible, Catechism, Torah, other)	1	2	3	4	5	6	7
i.	Websites on the Internet	1	2	3	4	5	6	7
j.	Research papers, reports, graphs, charts, or tables	1	2	3	4	5	6	7

Continue →

(5) Other than your regular English class, have you taken a class, <u>in school</u> this year intended to help you with your reading and writing?	Yes	No
	1❍ If **YES**, please continue to question 6	2❍ If **NO**, please continue to question 8

(6) 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❍	2❍	3❍	4❍	5❍

(7) 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❍	2❍	3❍	4❍	5❍	6❍	7❍

(8) Did an adult <u>in your school</u> help you individually with your reading and writing this year, like a tutor?	Yes	No
	1❍ If **YES**, please continue to question 9	2❍ If **NO**, please continue to question 11

(9) 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❍	2❍	3❍	4❍	5❍

(10) 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❍	2❍	3❍	4❍	5❍	6❍	7❍

Continue ➡

(11) 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
	1❍ If **YES**, please continue to question 12	2❍ If **NO**, please continue to question 14

(12) 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❍	2❍	3❍	4❍	5❍

(13) 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❍	2❍	3❍	4❍	5❍	6❍	7❍

(14) 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❍ If **YES**, please continue to question 15	2❍ If **NO**, please continue to question 17

(15) 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❍	2❍	3❍	4❍	5❍

(16) 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❍	2❍	3❍	4❍	5❍	6❍	7❍

Continue ➡

This section is about your classes in school this year.

(17) 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 English class.	1❍	2❍	3❍	4❍
b.	When I don't understand a word while reading for English class, I try to break the word down into smaller pieces.	1❍	2❍	3❍	4❍
c.	To help me understand what I'm reading for English class, I try to connect the things that are unfamiliar to me with things I already know.	1❍	2❍	3❍	4❍
d.	While reading for English class, I rarely make predictions about what will come next in a passage.	1❍	2❍	3❍	4❍
e.	When I'm reading for English class I stop once in a while and go over what I have read.	1❍	2❍	3❍	4❍
f.	When I don't know the meaning of a word while reading for English class, I often look at other words in the sentence or paragraph to help me understand.	1❍	2❍	3❍	4❍
g.	I try to identify whether what I'm reading for English class is fact or opinion.	1❍	2❍	3❍	4❍
h.	I rarely stop to summarize a passage or paragraph while reading for English class.	1❍	2❍	3❍	4❍

Continue ➡

(18) For which <u>one</u> of the following classes did you do the <u>most</u> reading during the past school year? Mark (✓) <u>one</u> answer.

1❍ History (or Social Studies) 2❍ Science 3❍ Math

(19) Please indicate how much you DISAGREE or AGREE with the following statements <u>about the class you chose in Question 18</u>. 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 class.	1❍	2❍	3❍	4❍
b. When I don't understand a word while reading for class, I try to break the word down into smaller pieces.	1❍	2❍	3❍	4❍
c. To help me understand what I'm reading for class, I try to connect the things that are unfamiliar to me with things I already know.	1❍	2❍	3❍	4❍
d. While reading for class, I rarely make predictions about what will come next in a passage.	1❍	2❍	3❍	4❍
e. When I'm reading for class I stop once in a while and go over what I have read.	1❍	2❍	3❍	4❍
f. When I don't know the meaning of a word while reading for class, I often look at other words in the sentence or paragraph to help me understand.	1❍	2❍	3❍	4❍
g. I try to identify whether what I'm reading for class is fact or opinion.	1❍	2❍	3❍	4❍
h. I rarely stop to summarize a passage or paragraph while reading for class.	1❍	2❍	3❍	4❍

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

(20) 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❍	2❍	3❍	4❍
b. Compared to work I do for other subjects at school, I find the work I do for ERO to be interesting.	1❍	2❍	3❍	4❍
c. Compared with what I learn in my other subjects at school, I find what I learn in ERO to be useful.	1❍	2❍	3❍	4❍

THANK YOU!!!

Appendix C

ERO Implementation Fidelity

This appendix describes the development of measures based on the classroom observation data collected during site visits to the Enhanced Reading Opportunities (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 spring 2006. The analysis of ERO program implementation fidelity in the second year of the study is based on field research visits to each of the 34 high schools during fall 2006 and spring 2007. 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 (RAAL) 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 for Year 1 of the study, 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 — had 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. A refresher training was provided before the Year 2 site visits to give the observers more practice using the protocols and to address any challenges that may have arisen during Year 1. The classroom observations in Year 1 were conducted by two researchers per school district (a senior staff member with at least a master's degree and a junior staff member with at least a bachelor's degree) and in Year 2 by one researcher per school district (a senior staff member with at least a master's degree). The observations captured between 160 and 180 minutes of instruction in each of the 34 high schools during each visit. 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).

[1] Biancarosa and Snow (2004).

Site visits were scheduled with the intent of observing classrooms across schools after similar amounts of instructional time had passed. On average, in the first year, the observations occurred 21 weeks after the ERO classes had 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. On average, in the second year, the fall observations occurred 15 weeks after the ERO classes had started and the spring observations took place about 16 weeks after the fall observations. The fall observations occurred at a point in time when teachers had gained some experience with the curriculum and with the teaching of the ERO programs. The spring observations occurred after the teachers had received their booster trainings with the developers and at a point when the teachers had covered much of the curriculum. The fact that the measurement of implementation fidelity in the second year of the study is based on two sets of classroom observations — as opposed to the first year of the study, when observations from only one site visit were used — also means that the fidelity measures in Year 2 capture a fuller range of teachers' experiences with the programs, which can be used to depict changes in implementation fidelity over the course of the school year.

During the visits to a given school, the observer took detailed field notes, focusing on teachers' presentation of curriculum components, the flow of instruction, students' behavior and engagement, and teacher-student interactions. The observer then gave a summative rating across all the observed classes in the school (ranging from two to four classes), 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 RAAL and Xtreme Reading). The 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 then 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.

---

[2]In 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 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.

- **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 observa-

tion process. Second, site visits were conducted by senior study team members.[3] In the second year, all of the observers had participated in the first year of site visits and were thoroughly trained on the observation instrument over the course of the two years of the study. Third, although a given observer conducted all observations in all of the participating high schools in a school district, the observers varied across districts, thus limiting the potential for the development of particularistic understandings by a given observer 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 observer(s) 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

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, for each of the two site visits. This section of the appendix describes how the composite measure of the learning environment dimension was calculated. The reliability of these constructs (Cronbach's alpha) is reported for both years. Because the reliability of these constructs is similar across the fall and spring site visits in the second year, the reliability for that year is reported for the spring site visit only.

Learning Environment Composite (2 items, Cronbach's alpha = .84 in Year 1, .86 in Year 2)

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.

[3]In the first year, at least one member of each pair of observers on a site visit was a senior study team member.

General Instructional Learning Environment Component (2 items, Cronbach's alpha = .77 in Year 1, .69 in Year 2)

This component is the average of two observed constructs that are part of the general instructional scales: classroom climate and on-task participation.[4]

Program-Specific Learning Environment Components

Reading Apprenticeship Academic Literacy (1 item, Cronbach's alpha = NA in both years)

The program-specific component of the learning environment composite for RAAL 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 in Year 1, .88 in Year 2)

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.

Equations B-1 and B-2 (below) show how the constructs and components were combined to calculate the learning environment composite measures for RAAL and Xtreme Reading schools.[5]

$$LE_{RA} = ½ (½ (GIC_1 + GIC_2) + (PSC_{RA1})) \quad \text{(B-1)}$$

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 (RAAL construct)

---

[4]In 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 in Appendix A, 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."

[5]In these equations, "LE" stands for learning environment, "RA" and "XR" stand for Reading Apprenticeship Academic Literacy and Xtreme Reading respectively, and "GIC" and "PSC" stand for general instructional construct and program-specific construct respectively.

$$LE_{XR} = ½ (½ (GIC_1 + GIC_2) + ½ (PSC_{XR1} + PSC_{XR2})) \quad \text{(B-2)}$$

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. As above, the reliability is reported for the spring site visit only.

Comprehension Instruction Composite (2 items, Cronbach's alpha = .72 in Year 1, .79 in Year 2)

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 schools and a program-specific component measured at each school — the RAAL component at each of the 17 RAAL 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 in Year 1, .70 in Year 2)

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 Academic Literacy (5 items, Cronbach's alpha = .70 in Year 1, .69 in Year 2)

The program-specific component of the comprehension instruction composite for RAAL 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 in Year 1, .63 in Year 2)

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 in Year 1, .69 in Year 2). The needs-driven instruction construct is the average of two subconstructs: student accommodations and feedback to students (Cronbach's alpha = .71 in Year 1, .51 in Year 2).

Equations B-3 and B-4 (below) show how the constructs and components were combined to calculate the comprehension instruction composite measures for RAAL and Xtreme Reading schools.[6]

$$CI_{RA} = \tfrac{1}{2}\left(\tfrac{1}{2}\left(GIC_1 + GIC_2\right) + 1/5\left(PSC_{RA1} + PSC_{RA2} + PSC_{RA3} + PSC_{RA4} + PSC_{RA5}\right)\right) \qquad \text{(B-3)}$$

Where:

CI_{RA} = comprehension instruction composite measure in a Reading Apprenticeship school
GI_{C1} = comprehension (general instructional construct)
GI_{C2} = metacognition (general instructional construct)
PSC_{RA1} = metacognitive conversations (RAAL construct)
PSC_{RA2} = silent sustained reading (RAAL construct)
PSC_{RA3} = content/theme integration (RAAL construct)
PSC_{RA4} = writing (RAAL construct)
PSC_{RA5} = integration of curriculum strands (RAAL construct)

$$CI_{XR} = \tfrac{1}{2}\left(\tfrac{1}{2}\left(GIC_1 + GIC_2\right) + \tfrac{1}{2}\left(PSC_{XR1} + PSC_{XR2}\right)\right) \qquad \text{(B-4)}$$

[6]In 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.

Where:

CI_{XR} = comprehension instruction composite measure in an Xtreme Reading school
GI_{C1} = comprehension (general instructional construct)
GI_{C2} = 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 average ratings calculated for each of the 34 participating high schools on the implementation fidelity of their classroom learning environment and on the implementation fidelity of their comprehension instruction. Each average 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.

The purpose of these fidelity groupings was to identify schools where the implementation of one or both of the two key program dimensions was problematic, and where programs were not an accurate representation of the program models. This was especially important in Year 1, when implementation of the programs in some of the schools was characterized by notable challenges.[7] Although program implementation was less problematic in Year 2 based on observer fidelity ratings from the two site visits, it remains important to identify schools whose programs are not aligned with developers' specifications, thus the fidelity groupings used in Year 2 are defined in the same way as in Year 1.

The range of average scores used to define each of the three fidelity groupings are described below ("well aligned," moderately aligned," and "poorly aligned"). Because the purpose of these groupings was to identify schools whose programs were not representative of the

[7]Please see Kemple et al. (2008) for a discussion of these challenges.

intended programs, also presented below is the number of constructs rated in Category 1 (the lowest score that can be assigned) in the set of schools that fell into the relevant grouping.

1. **Well aligned.** 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 turns out, the set of schools rated as well aligned in Year 2 had no more than one construct for each implementation dimension rated in Category 1 (the lowest score that can be assigned).

2. **Moderately aligned.** Implementation fidelity for the learning environment or comprehension instruction dimensions was characterized 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 the learning environment, the schools rated as moderately aligned in Year 2 had one construct rated in Category 1 (out of three or four constructs used to calculate the composite for RAAL 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.

3. **Poorly aligned.** 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. In schools rated as poorly aligned in Year 2, half or more of the general or program-specific constructs that make up the dimension were rated in Category 1. These programs were the least representative of the activities and practices intended by the respective program developers.

Based on their learning environment and comprehension ratings, schools were categorized into one of these three fidelity groupings. However, to prevent the risk of disclosure, the "moderately aligned" and "poorly aligned" categories are collapsed together in all report tables.

Appendix Tables C.1 and C.2 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 or moderately/poorly aligned categories of fidelity during the first two site visits, in spring of Year 1 and fall of Year 2, respectively. These tables are in the same format as Table 3.1 in Chapter 3, which presents implementation fidelity findings for the last site visit in spring of Year 2. Appendix Table C.3 presents the distribution of schools across these same categories of implementation fidelity, but based on the average of the fall and spring ratings for the learning environment and comprehension instruction dimensions in the second year. Appendix Tables C.4 and C.5 present average implementation composite scores by teachers' experience with the ERO program (that is, for the 25 schools where teachers taught two full years of the ERO program versus the nine schools where there were replacement teachers who taught less than two full years of the program), at each of the two site visits in Year 2.

The Enhanced Reading Opportunities Study

Appendix Table C.1

Number of ERO Classrooms with Well-, Moderately, or Poorly Aligned Implementation to Program Models on Each Implementation Dimension, by ERO Program: Year 1, Spring Site Visit

Implementation Dimension	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Learning environment			
Well-aligned implementation (composite rating is 2.0 or higher)	26	14	12
Moderately or poorly aligned implementation (composite rating is less than 2.0)	8	3	5
Comprehension instruction			
Well-aligned implementation (composite rating is 2.0 or higher)	16	7	9
Moderately or poorly aligned implementation (composite rating is less than 2.0)	18	10	8
Combined dimensions			
Well-aligned implementation on both dimensions	16	7	9
Moderately or poorly aligned implementation on at least one dimension	18	10	8
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 begninning 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.

The Enhanced Reading Opportunities Study

Appendix Table C.2

Number of ERO Classrooms with Well-, Moderately, or Poorly Aligned Implementation to Program Models on Each Implementation Dimension, by ERO Program: Year 2, Fall Site Visit

Implementation Dimension	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Learning environment			
School average	2.46	2.47	2.44
Well-aligned implementation (composite rating is 2.0 or higher)	30	--	--
Moderately or poorly aligned implementation (composite rating is less than 2.0)	4	--	--
Comprehension instruction			
School average	2.10	2.10	2.10
Well-aligned implementation (composite rating is 2.0 or higher)	20	11	9
Moderately or poorly aligned implementation (composite rating is less than 2.0)	14	6	8
Combined dimensions			
School average	2.28	2.28	2.27
Well-aligned implementation on both dimensions	20	11	9
Moderately or poorly aligned implementation on at least one dimension	14	6	8
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 begninning 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.

Some counts have been suppressed to protect confidentiality.

The Enhanced Reading Opportunities Study

Appendix Table C.3

Number of ERO Classrooms with Well-, Moderately, or Poorly Aligned Implementation to Program Models on Each Implementation Dimension, by ERO Program: Year 2, Fall and Spring Site Visits

Implementation Dimension	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Learning environment			
School average	2.46	2.55	2.36
Well-aligned implementation (composite rating is 2.0 or higher)	31	--	--
Moderately or poorly aligned implementation (composite rating is less than 2.0)	3	--	--
Comprehension instruction			
School average	2.21	2.18	2.24
Well-aligned implementation (composite rating is 2.0 or higher)	23	11	12
Moderately or poorly aligned implementation (composite rating is less than 2.0)	11	6	5
Combined dimensions			
School average	2.33	2.37	2.30
Well-aligned implementation on both dimensions	23	11	12
Moderately or poorly aligned implementation on at least one dimension	11	6	5
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 begninning 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.

Some counts have been suppressed to protect confidentiality.

The Enhanced Reading Opportunities Study

Appendix Table C.4

Average Implementation Composite Scores, by Experience with the ERO Program: Year 2, Fall

Characteristic	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Learning environment			
Schools with replacement teachers	2.5	2.6	2.5
Schools with teachers who taught two full years	2.4	2.4	2.4
Comprehension instruction			
Schools with replacement teachers	2.2	2.2	2.1
Schools with teachers who taught two full years	2.1	2.1	2.1
Combined dimensions			
Schools with replacement teachers	2.4	2.4	2.3
Schools with teachers who taught two full years	2.2	2.3	2.2

SOURCES: MDRC and AIR calculations from classroom observation data.

NOTES: There were 25 teachers who taught all of Year 1, 13 at Reading Apprenticeship schools and 12 at Xtreme Reading schools. There were 9 replacement teachers, 4 at Reading Apprenticeship schools and 5 at Xtreme Reading schools.

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 begninning 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.

The Enhanced Reading Opportunities Study

Appendix Table C.5

Average Implementation Composite Scores, by Experience with the ERO Program: Year 2, Spring

Characteristic	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Learning environment			
Schools with replacement teachers	2.4	2.6	2.3
Schools with teachers who taught two full years	2.5	2.6	2.3
Comprehension instruction			
Schools with replacement teachers	2.3	2.3	2.2
Schools with teachers who taught two full years	2.3	2.3	2.4
Combined dimensions			
Schools with replacement teachers	2.3	2.5	2.2
Schools with teachers who taught two full years	2.4	2.4	2.4

SOURCES: MDRC and AIR calculations from classroom observation data.

NOTES: There were 25 teachers who taught all of Year 1, 13 at Reading Apprenticeship schools and 12 at Xtreme Reading schools. There were 9 replacement teachers, 4 at Reading Apprenticeship schools and 5 at Xtreme Reading schools.

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 begninning 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.

Classroom Observation Scales

Enhanced Reading Opportunities Study

American Institutes for Research

November 2006

Table of Contents

General Instruction Scales 1
RAAL Fidelity Measure 6
Xtreme Reading Fidelity Measure 14

Enhanced Reading Opportunities Program

General Instruction Scales

Area of interest	Basic Literacy Skills (Advanced phonics and decoding, fluency)

Description

0. Not applicable. During the observed class period(s), students do not demonstrate a need for instruction in basic literacy skills.*
1. During the observed class period(s), instruction does not reflect teacher recognition of a *demonstrated* 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).
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.
3. During the observed class period(s), instruction reflects teacher recognition of student difficulty with basic literacy skills *and* 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).

*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

0. Not applicable. There was no opportunity for vocabulary instruction to occur during the observed class period(s).
1. 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.
2. Students are engaged in some vocabulary activities, but these activities are not fully developed. For example, the teacher may employing 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).
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

1. 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.
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.
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.

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.)

1. Little metacognitive work is apparent, and overall, metacognitive skills are not being *developed* 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").
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").
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	

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.
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.
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	

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.
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.
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.

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-

* 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.

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

0. SSR did not take place during the observed class period(s).

1. 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

0. Not applicable. Language Study did not take place during the observed class period(s).

1. 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

0. Not applicable. The observed class period(s) did not include a writing component.

1. 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)

0. There was no opportunity to make accommodations during the observed class period(s).
1. 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)

0. There is no opportunity to provide feedback to students during the observed period (s).
1. There is opportunity but 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 D

State Tests Included in the ERO Study

This appendix discusses various issues related to the use of state tests in the impact analyses presented in this report. As noted in Chapter 2, there is variation across the study districts in terms of the timing, content, and scale of their state tests. This variation complicates the analysis of these scores and leads to uncertainty as to how to interpret the state test findings. To provide further detail on these issues, the first section of this appendix describes the tests included in the state test analysis and the number of students with scores for each of the tests. The second section describes the methods used to standardize the scores on these tests in order to pool the data across districts for the analysis.

Characteristics of State Tests

Appendix Table D.1 describes the assessments included in the impact analyses, by core subject area (English language arts [ELA], social studies, science, and math) and by year (students' ninth- and tenth- grade year, which generally correspond to the program year and follow-up year, respectively). The table lists the name of each assessment as well as the number of students in the school records sample who have a test score on the assessment.[1] The table also indicates which of these tests are used in establishing a school's Adequate Yearly Progress (AYP) as defined by the No Child Left Behind act (NCLB). As seen in this table:

- Not all districts administer tests in every subject area in any given year.

Also, within districts, not all students take the test in the same school year. Many of the assessments are "end-of-course" tests which students take upon finishing a specific class, and therefore, there is variation across students in terms of when they write the assessment. For these reasons, for any given subject area and in any given year, less than 50 percent of students in the full study sample have a test score.

- There is variation in the districts' purposes for administering each test.

Most, but not all, of the tests are used for some type of student accountability, either as part of a course grade or as a prerequisite for graduation (some of these tests are also used for school accountability). Two of the tests are used solely for school accountability at the district, state, or federal level. These are the Texas Assessment of Knowledge and Skills (TAKS)

[1] In cases where students took a given test more than once, only the first test score is used. In cases where students took more than one test per subject area during the school year (for example, if a student took two science tests, one in biology and one in physics), their score on the test taken by the most students in the district that year is used.

The Enhanced Reading Opportunities Study

Appendix Table D.1

Description of State Tests, by District

English Language Arts (ELA)

District	Ninth Grade	Number of Students[a]	Tenth Grade	Number of Students[a]
1			Maryland High School Assessment: English[b]	450
2	Georgia End-of-Course Tests: Ninth Grade Literature and Composition	531	Georgia End-of-Course Tests: Ninth Grade Literature and Composition	34
			Georgia End-of-Course Tests: American Literature and Composition	11
3			University of New York State Regents High School Examination: Comprehensive Examination in English[b]	4
4	Texas Assessment of Knowledge and Skills (TAKS): Grade 9 Reading	520	Texas Assessment of Knowledge and Skills (TAKS): Grade 10 Reading[b]	394
5			Virginia Standards of Learning: End-of-course English: Reading, Literature and Research[b]	18
6	Utah Performance Assessment System for Students: English Language Arts Criterion-Referenced Test Grade 9	300	Utah Performance Assessment System for Students: English Language Arts Criterion-Referenced Test Grade 10[b]	249
7			Nebraska School Based Teacher-led Assessment and Reporting System (STARS): PLAN English[b]	278
8	South Carolina End-of-Course-Examination Program (EOCEP): English 1	230	South Carolina High School Assessment Program (HSAP): English Language Arts[b]	240
9	Texas Assessment of Knowledge and Skills (TAKS): Grade 9 Reading	663	Texas Assessment of Knowledge and Skills (TAKS): Grade 10 Reading[b]	583
10			University of New York State Regents High School Examination: Comprehensive Examination in English[b]	147
Total		2,244		2,408

(continued)

Appendix Table D.1 (continued)

Social Studies

District	Ninth Grade	Number of Students[a]	Tenth Grade	Number of Students[a]
1			Maryland High School Assessment: Government	450
2	Georgia End-of-Course Tests: United States History	6	Georgia End-of-Course Tests: United States History	8
3	University of New York State Regents High School Examination: Global History and Geography	4	University of New York State Regents High School Examination: Global History and Geography	244
4	Stanford Achievement Test, Tenth Edition (SAT 10): Social Science, Grade 9[c]	502	Texas Assessment of Knowledge and Skills (TAKS): Grade 10 Social Studies	290
5			Virginia Standards of Learning: Virginia and United States History	113
	Virginia Standards of Learning: World Geography	24	Virginia Standards of Learning: World Geography	3
	Virginia Standards of Learning: World History and Geography to 1500 A.D.	303	Virginia Standards of Learning: World History and Geography to 1500 A.D.	6
	Virginia Standards of Learning: World History and Geography 1500 A.D. to Present	113	Virginia Standards of Learning: World History and Geography 1500 A.D. to Present	221
6				
7				
8				
9			Texas Assessment of Knowledge and Skills (TAKS): Grade 10 Social Studies	501
10			University of New York State Regents High School Examination: Global History and Geography	385
			University of New York State Regents High School Examination: United States History and Government	16
Total		952		2,237

(continued)

Appendix Table D.1 (continued)

Science

District	Ninth Grade	Number of Students[a]	Tenth Grade	Number of Students[a]
1	Maryland High School Assessment: Biology	39	Maryland High School Assessment: Biology	393
2	Georgia End-of-Course Tests: Biology	208	Georgia End-of-Course Tests: Biology	246
	Georgia End-of-Course Tests: Physical Science	273	Georgia End-of-Course Tests: Physical Science	34
3	University of New York State Regents High School Examination: Physical Setting Earth Science	71	University of New York State Regents High School Examination: Physical Setting Earth Science	117
	University of New York State Regents High School Examination: Living Environments	164	University of New York State Regents High School Examination: Living Environments	84
	University of New York State Regents High School Competency Test: Science (for students with IEPs)	5		
4	Stanford Achievement Test, Tenth Edition (SAT 10): Science, Grade 9[c]	500	Texas Assessment of Knowledge and Skills (TAKS): Grade 10 Science	290
5	Virginia Standards of Learning: End-of-course Earth Science	263	Virginia Standards of Learning: End-of-course Earth Science	73
	Virginia Standards of Learning: End-of-course Biology	114	Virginia Standards of Learning: End-of-course Biology	221
			Virginia Standards of Learning: End-of-course Chemistry	10
6	Utah Performance Assessment System for Students: Earth Systems Science Criterion-Referenced Test	227	Utah Performance Assessment System for Students: Biology Criterion-Referenced Test	78
	Utah Performance Assessment System for Students: Biology Criterion-Referenced Test	9	Utah Performance Assessment System for Students: Physics Criterion-Referenced Test	17
7			Nebraska School Based Teacher-led Assessment and Reporting System (STARS): PLAN Science	275
8	South Carolina End-of-Course-Examination Program (EOCEP): Physical Science	118		
9			Texas Assessment of Knowledge and Skills (TAKS): Grade 10 Science	499
10	University of New York State Regents High School Examination: Physical Setting Earth Science	68	University of New York State Regents High School Examination: Physical Setting Earth Science	64
	University of New York State Regents High School Examination: Living Environments	289	University of New York State Regents High School Examination: Living Environments	260
Total		2,348		2,661

(continued)

Appendix Table D.1 (continued)

Math

District	Ninth Grade	Number of Students[a]	Tenth Grade	Number of Students[a]
1	High School Assessment: Algebra/Data Analysis[b]	478	High School Assessment: Algebra/Data Analysis[b]	199
2	Georgia End-of-Course Tests: Algebra 1	362	Georgia End-of-Course Tests: Algebra 1	42
	Georgia End-of-Course Tests: Geometry	116	Georgia End-of-Course Tests: Geometry	252
3	University of New York State Regents High School Examination: Mathematics A[b]	4	University of New York State Regents High School Examination: Mathematics A[b]	165
	University of New York State Regents High School Competency Test: Math (for students with IEPs)	9	University of New York State Regents High School Examination: Integrated Algebra[b]	18
4	Texas Assessment of Knowledge and Skills (TAKS): Grade 9 Mathematics	512	Texas Assessment of Knowledge and Skills (TAKS): Grade 10 Mathematics[b]	385
5	Virginia Standards of Learning: Algebra I[b]	253	Virginia Standards of Learning: Algebra I[b]	103
	Virginia Standards of Learning: Geometry[b]	41	Virginia Standards of Learning: Algebra II[b]	27
			Virginia Standards of Learning: Geometry[b]	125
6				
7			Nebraska School Based Teacher-led Assessment and Reporting System (STARS): PLAN Mathematics[b]	277
8	South Carolina End-of-Course-Examination Program (EOCEP): Algebra 1	169		
9	Texas Assessment of Knowledge and Skills (TAKS): Grade 9 Mathematics	663	Texas Assessment of Knowledge and Skills (TAKS): Grade 10 Mathematics[a]	578
10	University of New York State Regents High School Examination: Mathematics A[b]	61	University of New York State Regents High School Examination: Mathematics A[b]	299
			University of New York State Regents High School Examination: Integrated Algebra[b]	48
			University of New York State Regents High School Examination: Mathematics B[b]	19
Total		2,668		2,537

NOTES: [a]The "Number of Students" refers to the number of students included in the state test analysis, which consists of students in the school records sample who have valid test scores for the given year. If a student took a given test more than once, only his or her first score is included in the analysis. If a student took more than one test per subject during the same year, his or her score on the test taken by the most students in the district that year is used.

[b]This test is used to determine Adequate Yearly Progress (AYP) under the No Child Left Behind Act of 2001 (Pub. L. 107-110).

[c]This test is administered across Houston Public Schools District but is not a state requirement.

administered in two districts,[2] and the Stanford Achievement Tests, Tenth Edition (SAT 10) administered in one district.[3]

- There is also substantial variation in the content of the different tests.

Some districts administer general achievement tests while others administer tests that are course-specific. For example, science assessments used in the analysis for this report include tests in general science, biology, living environments, earth science, physical science, and chemistry.

Standardization of State Test Scores

The state tests described above also vary in terms of the scale of their scores. In order to pool the data from these different tests, state test scores were therefore converted to a common metric. Specifically, scaled scores for each test were standardized (z-scored) by district, by year (program year or follow-up year), and by cohort, based on the mean and standard deviation of the non-ERO group, as follows:[4,5]

$$Z_{icjt} = \frac{(Y_{icjt} - \bar{Y}_{cjt})}{\sigma_{cjt}}$$

where:

Z_{icjt} = the standardized score for student *i* (in cohort *c* from site *j*) who took the test in year *t*.

Y_{icjt} = the scaled (raw) score for student *i* (in cohort *c* from site *j*) who took the test in year *t*.

---

[2]The exit-level version of the TAKS, which is administered in later grades (and therefore not captured in this study), is used for student accountability.

[3]The SAT 10 tests in Social Science and Science, which are administered throughout the Houston Public Schools district, are the only tests in the analysis that are not state-required assessments.

[4]Scores are standardized by district because the test varies by district. Scores are also standardized by cohort and year (program year or follow-up year), because a test's content and scale may differ across school years.

[5]As explained in May et al. (2009), impact estimates can be combined across states or districts by converting test scores to z-scores. When calculating z-scores, it is appropriate to use the mean and standard deviation of the sample (or in this case the non-ERO group) when students from each district/state represent a similar cross-section of the population of students targeted by the intervention. This assumption is met in the ERO study, given the way in which students were recruited (students in each district are all two to five years below grade level).

$\bar{Y}_{cjt}$ = the average scaled (raw) score for non-ERO students in cohort *c* from site *j* who took the test in year *t*.

σ_{cjt} = the standard deviation of scaled (raw) scores for students in cohort *c* from site *j* who took the test in year *t*.

This z-score transformation converts the scaled scores for each test to a common metric, the *effect size unit* (or standard deviation unit). Standardized scores represent the amount by which a student's score is above or below average relative to the typical student in the counterfactual condition (as represented by the average score among non-ERO students), as a proportion of the total variation in test scores in the counterfactual condition (non-ERO students). For example, if a student has a z-score of 0.20 on a test, this means that his or her test score surpasses that of the average student in the non-ERO group by 20 percent of the standard deviation in test scores for the non-ERO group.[6]

Impact analyses were conducted using the standardized scores (z-scores) as the outcome measure. Given the metric of the transformed test scores, the estimated impact of the ERO programs on students' performance on state test scores is scaled as an effect size.

[6]In this report, the mean and standard deviation of the non-ERO group are used as the reference point when standardizing test scores, because the non-ERO group represents the counterfactual condition in the impact analysis. More broadly, in this report, effect sizes for impacts on all high school outcomes (grade point average, credits earned, school behaviors) are always calculated based on the standard deviation of the non-ERO group (see discussion of effect sizes in Chapter 2 and the impact table in Chapter 4).

Appendix E

Response Analysis and Baseline Comparison Tables

This appendix serves as a supplement to the description of students presented in Chapter 2. In particular, additional information is presented on the baseline characteristics of students in the full study sample, on the availability of school records data among study participants, and on the baseline characteristics of students included in the impact analyses (analysis samples).

In most of the tables in this appendix, the baseline characteristics and achievement of two groups of students are compared (whether the Enhanced Reading Opportunities [ERO] group and the non-ERO group, or students in the analysis sample and students excluded from the sample). Because many hypothesis tests are conducted in these tables (one for each student baseline characteristic), there is an increased probability of concluding that a particular baseline difference is statistically significant when in fact it is not (this is called a Type I error or a "false positive").[1] For this reason, we use an omnibus test to test for a systematic or overall difference between the characteristics of the ERO and non-ERO group. This test is reported at the bottom of each of the tables in this appendix. If the omnibus test is not statistically significant, this means that a statistically significant difference for any given baseline characteristic in the tables may be due to chance.

Baseline Characteristics of the Full Study Sample

The first set of tables compares the baseline characteristics of ERO and non-ERO students in the full study sample (that is, all randomly assigned students who were two to five years behind grade level at the time of baseline testing). Appendix Table E.1 presents this comparison for all schools in the study, while Appendix Tables E.2 and E.3 display this comparison for each ERO program separately (Reading Apprenticeship Academic Literacy [RAAL] schools and Xtreme Reading schools):

- In the full study sample, students in the ERO and non-ERO groups do not systematically differ from each other in terms of their baseline characteristics.

Although there are statistically significant differences between the two program groups on some of the individual characteristics in the tables (for example, baseline reading comprehension scores), a joint test of the difference between the two groups on all characteristics is not statistically significant, whether overall or by program. The lack of a systematic difference indicates that random assignment was successful in creating two equivalent research groups at baseline.

[1] In particular, one would expect to see a "false positive" for every 20 hypothesis tests conducted.

The Enhanced Reading Opportunities Study

Appendix Table E.1

Characteristics of the Full Study Sample

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.744
Hispanic	31.1	30.9	0.2	
Black, non-Hispanic	46.1	46.9	-0.7	
White, non-Hispanic	16.6	15.8	0.8	
Other	6.1	6.4	-0.3	
Male (%)	49.7	51.5	-1.8	0.180
Average age (years)	14.8	14.8	0.0	0.399
Overage for grade[b] (%)	30.2	29.3	0.9	0.446
Language other than English spoken at home (%)	46.8	47.6	-0.8	0.499
Mother's education level (%)				0.575
Did not finish high school	19.3	19.3	0.0	
High school diploma or GED certificate	25.8	26.2	-0.3	
Completed some postsecondary education	31.5	32.8	-1.3	
Don't know	23.3	21.7	1.6	
GRADE reading comprehension[c]				
Average standard score	85.1	85.4	-0.3 *	0.045
Corresponding grade equivalent	*5.0*	*5.1*		
Corresponding percentile	*15*	*16*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	1.98	1.94	0.04	0.121
English language arts (ELA)	2.02	2.00	0.02	0.452
Math	1.89	1.84	0.05	0.094
Social studies	2.06	2.01	0.04	0.136
Science	1.96	1.92	0.05	0.109
Credits earned in core subjects (as a percentage of credits attempted)	87.6	86.5	1.1	0.099
Attendance rate (%)	94.7	94.5	0.2	0.247
Free and reduced-price lunch (%)	67.1	67.1	0.0	0.983
Joint test of difference between program groups[e] ($\chi 2 = 42.7$)				0.239
Sample size[f]	3,204	2,391		

(continued)

Appendix Table E.1 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of students.

The Enhanced Reading Opportunities Study

Appendix Table E.2

Characteristics of the Full Study Sample, Reading Apprenticeship Schools

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.592
Hispanic	30.1	30.1	0.0	
Black, non-Hispanic	45.9	47.6	-1.7	
White, non-Hispanic	17.1	16.3	0.9	
Other	6.8	6.0	0.8	
Male (%)	49.6	51.4	-1.8	0.352
Average age (years)	14.8	14.8	0.0	0.900
Overage for grade[b] (%)	30.2	30.0	0.1	0.936
Language other than English spoken at home (%)	44.9	47.3	-2.4	0.161
Mother's education level (%)				0.997
Did not finish high school	20.2	19.7	0.5	
High school diploma or GED certificate	26.5	26.5	0.0	
Completed some postsecondary education	30.2	30.5	-0.2	
Don't know	23.1	23.3	-0.2	
GRADE reading comprehension[c]				
Average standard score	85.1	85.2	-0.1	0.664
Corresponding grade equivalent	*5.0*	*5.0*		
Corresponding percentile	*15*	*15*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	1.98	1.93	0.05	0.124
English language arts (ELA)	2.04	2.00	0.04	0.380
Math	1.90	1.83	0.06	0.146
Social studies	2.02	1.97	0.05	0.227
Science	1.99	1.93	0.05	0.208
Credits earned in core subjects (as a percentage of credits attempted)	87.6	86.1	1.4	0.129
Attendance rate (%)	94.4	94.2	0.2	0.490
Free and reduced-price lunch (%)	67.9	65.7	2.2	0.178
Joint test of difference between program groups[e] ($\chi 2 = 27.1$)				0.883
Sample size[f]	1,594	1,168		

(continued)

Appendix Table E.2 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of students.

The Enhanced Reading Opportunities Study

Appendix Table E.3

Characteristics of the Full Study Sample, Xtreme Reading Schools

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.499
Hispanic	32.0	31.7	0.4	
Black, non-Hispanic	46.3	46.1	0.2	
White, non-Hispanic	16.1	15.4	0.7	
Other	5.5	6.8	-1.3	
Male (%)	49.7	51.5	-1.8	0.335
Average age (years)	14.8	14.8	0.0	0.288
Overage for grade[b] (%)	30.2	28.6	1.6	0.325
Language other than English spoken at home (%)	48.7	48.0	0.7	0.660
Mother's education level (%)				0.186
Did not finish high school	18.4	18.9	-0.5	
High school diploma or GED certificate	25.2	25.8	-0.6	
Completed some postsecondary education	32.8	35.1	-2.3	
Don't know	23.6	20.1	3.5*	
GRADE reading comprehension[c]				
Average standard score	85.1	85.6	-0.5*	0.016
Corresponding grade equivalent	*5.0*	*5.1*		
Corresponding percentile	*15*	*16*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	1.97	1.95	0.02	0.515
English language arts (ELA)	2.01	2.00	0.01	0.860
Math	1.89	1.85	0.04	0.365
Social studies	2.09	2.06	0.04	0.370
Science	1.94	1.90	0.04	0.313
Credits earned in core subjects (as a percentage of credits attempted)	87.7	86.9	0.7	0.417
Attendance rate (%)	94.9	94.7	0.2	0.342
Free and reduced-price lunch (%)	66.4	68.6	-2.2	0.180
Joint test of difference between program groups[e] ($\chi 2 = 47.0$)				0.125
Sample size[f]	1,610	1,223		

(continued)

Appendix Table E.3 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of students.

Availability of Data

As explained in Chapter 2, data are not consistently available for all of the study outcomes. Appendix Tables E.4 and E.5 present the percentage of students in the full study sample for whom different types of data are available, for each reading program separately (RAAL schools and Xtreme Reading schools). The findings in these tables mirror those presented in Chapter 2 for all schools in the study. That is:[2]

- Students in the study sample are more likely to have information on their performance in core courses (grade point average [GPA] and credit accumulation) than on other outcomes. This is because if a student is enrolled at the school for at least one semester during the school year, the school district will have course transcript information for that student. That said, the availability of information on course performance decreases in the follow-up year. These findings hold for both RAAL schools and Xtreme Reading schools.

- The ERO and non-ERO group do not differ by a statistically significant amount with respect to the percentage of students who have data on high school outcomes (GPA, credits earned, state tests, and school behaviors). This result applies to both RAAL schools and Xtreme Reading schools.

- However, in RAAL schools, there is a statistically significant difference between the percentage of ERO and non-ERO students who completed the reading assessment and survey at the end of the program year. In schools implementing the RAAL program, ninth-grade testing and survey data are available for 84 percent of students in the ERO group, while these data are available for 79 percent of students in the non-ERO group; this difference of 4 percentage points is statistically significant. In Xtreme Reading schools, however, the between-group difference in response rates for the reading test and survey is not statistically significant.

Baseline Characteristics of the School Records Sample

As described in Chapter 2, impacts on high school outcomes (presented in Chapter 4 of this report) are based on students who, at minimum, have course performance data on GPA and

[2]These tables are the program-specific counterparts to Table 2.4, which presents this information for all schools in the study.

The Enhanced Reading Opportunities Study

Appendix Table E.4

Availability of Data for the Full Study Sample, by Data Source and Type of Outcome, Reading Apprenticeship Schools

Measure	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
GRADE reading assessment				
Program year (%)	83.5	79.4	4.1*	0.006
Student survey				
Program year (%)	83.5	79.1	4.4*	0.003
School records				
GPA and credit accumulation				
Program year (%)	92.1	93.7	-1.6	0.100
Follow-up year (%)	80.6	79.3	1.3	0.382
Performance on state tests				
Program year (%)				
English language arts	37.8	37.3	0.6	0.587
Social studies	16.4	16.3	0.1	0.858
Science	41.7	41.7	0.0	0.986
Math	45.5	45.6	-0.1	0.956
Follow-up year (%)				
English language arts	42.2	40.7	1.5	0.277
Social studies	40.1	39.8	0.3	0.827
Science	46.3	46.3	0.0	0.987
Math	45.0	43.7	1.3	0.432
School behaviors				
Program year (%)				
Attendance rate	90.9	92.3	-1.4	0.171
Ever suspended	84.2	85.6	-1.4	0.104
Follow-up year (%)				
Attendance rate	79.6	78.5	1.2	0.455
Ever suspended	73.5	71.9	1.5	0.292
Sample size[a]	1,594	1,168		

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

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]The sample sizes reported here are for the full study sample.

The Enhanced Reading Opportunities Study

Appendix Table E.5

Availability of Data for the Full Study Sample, by Data Source and Type of Outcome, Xtreme Reading Schools

Measure	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
GRADE reading assessment				
Program year (%)	83.3	80.9	2.4	0.090
Student survey				
Program year (%)	83.2	80.4	2.8	0.052
School records				
GPA and credit accumulation				
Program year (%)	91.2	91.4	-0.1	0.894
Follow-up year (%)	78.1	79.2	-1.1	0.480
Performance on state tests				
Program year (%)				
English language arts	41.7	41.7	0.0	0.971
Social studies	18.5	18.2	0.3	0.676
Science	41.6	43.5	-1.9	0.174
Math	50.2	48.6	1.6	0.239
Follow-up year (%)				
English language arts	44.8	44.8	0.0	0.986
Social studies	41.7	40.5	1.2	0.409
Science	50.4	48.3	2.1	0.219
Math	47.0	45.9	1.1	0.491
School behaviors				
Program year (%)				
Attendance rate	90.6	91.2	-0.6	0.575
Ever suspended	79.6	78.9	0.7	0.429
Follow-up year (%)				
Attendance rate	77.5	79.0	-1.6	0.315
Ever suspended	67.4	68.0	-0.7	0.648
Sample size[a]	1,610	1,223		

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

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]The sample sizes reported here are for the full study sample.

credit accumulation in a given year. Appendix Figure E.1 depicts the construction of this "school records sample" in the program year and the follow-up year.

Comparison of Students Included and Excluded from the School Records Sample

In order to understand the generalizability of the impact findings in this report, Appendix Tables E.6 through E.8 compare the baseline characteristics of students in the school records sample (program year and follow-up year) with the characteristics of students excluded from these analysis samples due to missing data. Tables E.6a and E.6b present this comparison for all schools in the study, while Tables E.7a through E.8b present this comparison for each reading program separately (RAAL schools and Xtreme Reading schools) in the program and follow-up year. These tables show that:

- There are systematic baseline differences between students with and without school records data. Students without school records data are more likely to have characteristics associated with a risk of school failure.

Specifically, students without school records data had lower reading comprehension test scores at baseline than students with school records data, and are more likely to be overage for the ninth grade. Additionally, these students had lower GPAs in core subjects, earned fewer core credits, and attended school less in the school year prior to ERO. In the follow-up year, differences between students with and without school records data become more pronounced. These differences indicate that the impact findings in this report may not be generalizable to students who are excluded from the analysis because they do not have school records data. However, because the analysis sample includes most students in the study sample (92 percent), the findings in this report are likely to still be generalize to the full study sample.

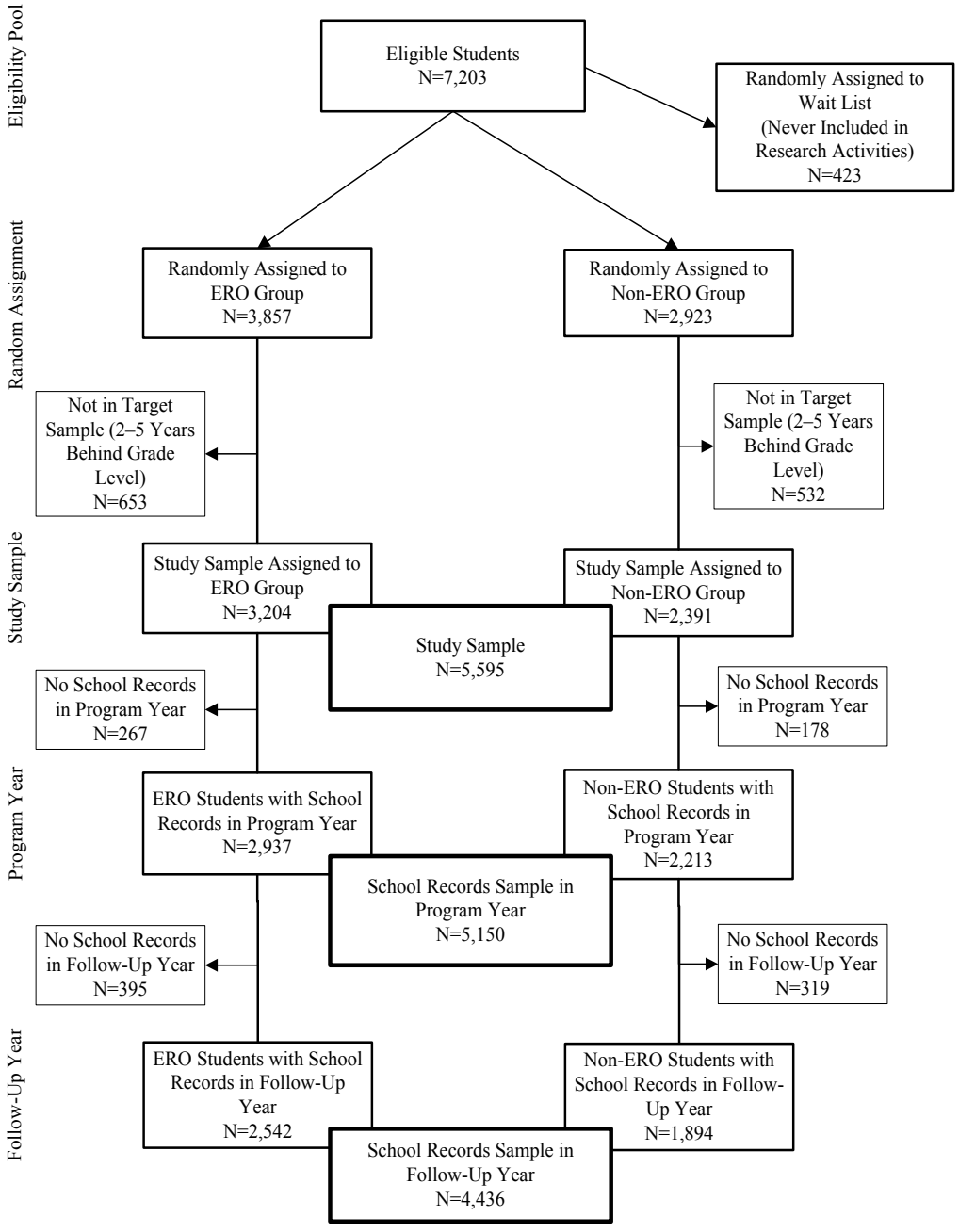
As explained in Chapter 2, however, data on state test scores are not consistently available for all students in the school records sample, due to variation in the timing of test-taking across the study districts. Thus, impacts on state tests are estimated based on different subsamples of students in the school records sample, which vary by subject area and year. Appendix Tables E.9 through E.11 compare the baseline characteristics of students with state test data in each subject area during each year with the characteristics of students in the full study sample, for all schools and by ERO program. These tables indicate that:

- Students with state test data in a given subject area and year are statistically different at baseline from students in the full study sample without the corresponding state test data.

The Enhanced Reading Opportunities Study

Appendix Figure E.1

Creation of the Primary Analysis Sample



The Enhanced Reading Opportunities Study

Appendix Table E.6a

Characteristics of Students:
Differences Between Students With and Without School Records in the Program Year

Characteristic	School Records Sample	No School Records Data	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)			*	0.000
Hispanic	31.9	21.4	10.5	
Black, non-Hispanic	46.4	51.3	-4.9	
White, non-Hispanic	15.4	22.5	-7.1	
Other	6.3	4.8	1.5	
Male (%)	50.4	50.7	-0.3	0.912
Average age (years)	14.8	15.0	-0.2*	0.000
Overage for grade[b] (%)	28.8	43.0	-14.2*	0.000
Language other than English spoken at home (%)	47.6	42.8	4.7*	0.047
Mother's education level (%)			*	0.000
Did not finish high school	19.0	22.2	-3.2	
High school diploma or GED certificate	25.7	29.5	-3.8	
Completed some postsecondary education	32.3	29.8	2.5	
Don't know	23.0	18.5	4.4	
GRADE reading comprehension[c]				
Average standard score	85.3	84.2	1.1*	0.000
Corresponding grade equivalent	*5.0*	*4.8*		
Corresponding percentile	*15*	*13*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)				
English language arts (ELA)	2.0	1.7	0.3*	0.000
Math	1.9	1.5	0.4*	0.000
Social studies	2.1	1.7	0.4*	0.000
Science	2.0	1.7	0.2*	0.000
Credits earned in core subjects (as a percentage of credits attempted)	87.7	80.3	7.4*	0.000
Attendance rate (%)	94.7	92.7	2.0*	0.000
Free and reduced-price lunch (%)	68.3	44.8	23.5*	0.000
Joint test of difference between program groups[e] ($\chi 2 = 751.9$)			*	0.000
Sample size[f]	5,150	445		

(continued)

Appendix Table E.6a (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2007-208 school year for Cohort 2. Estimated differences between the school records sample and students without school records data are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "School Records Sample" are the observed means for students in the school records sample. The values in the "No School Records Data" column are the regression-adjusted means for students without school records data, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between students in the school records sample and students missing school records data. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference at baseline between students with or without school records data, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full study sample.

The Enhanced Reading Opportunities Study

Appendix Table E.6b

Characteristics of Students: Differences Between Students With and Without School Records in the Follow-Up Year

Characteristic	School Records Sample	No School Records Data	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.336
Hispanic	31.3	30.1	1.3	
Black, non-Hispanic	46.3	48.6	-2.3	
White, non-Hispanic	15.9	15.9	0.0	
Other	6.4	5.4	1.0	
Male (%)	50.0	52.2	-2.3	0.231
Average age (years)	14.7	15.0	-0.3*	0.000
Overage for grade[b] (%)	25.8	45.5	-19.7*	0.000
Language other than English spoken at home (%)	47.3	46.8	0.5	0.779
Mother's education level (%)			*	0.000
Did not finish high school	18.1	23.8	-5.7	
High school diploma or GED certificate	26.5	24.1	2.4	
Completed some postsecondary education	32.3	31.4	0.9	
Don't know	23.1	20.7	2.4	
GRADE reading comprehension[c]				
Average standard score	85.4	84.6	0.8*	0.000
Corresponding grade equivalent	*5.1*	*4.9*		
Corresponding percentile	*16*	*14*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)				
English language arts (ELA)	2.1	1.8	0.3*	0.000
Math	1.9	1.6	0.3*	0.000
Social studies	2.1	1.8	0.3*	0.000
Science	2.0	1.7	0.3*	0.000
Credits earned in core subjects (as a percentage of credits attempted)	88.6	81.0	7.6*	0.000
Attendance rate (%)	95.2	91.9	3.3*	0.000
Free and reduced-price lunch (%)	68.5	60.9	7.7*	0.007
Joint test of difference between program groups[e] ($\chi 2 = 779.6$)			*	0.000
Sample size[f]	4,436	1,159		

(continued)

Appendix Table E.6b (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The follow-up year is the school year following student's enrollment in the ERO class; it corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2. Estimated differences between the school records sample and students without school records data are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "School Records Sample" are the observed means for students in the school records sample. The values in the "No School Records Data" column are the regression-adjusted means for students without school records data, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between students in the school records sample and students missing school records data. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference at baseline between students with or without school records data, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full study sample.

The Enhanced Reading Opportunities Study

Appendix Table E.7a

Characteristics of Students:
Differences Between Students With and Without School Records in the Program Year,
Reading Apprenticeship Schools

Characteristic	School Records Sample	No School Records Data	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)			*	0.000
Hispanic	31.3	19.8	11.5	
Black, non-Hispanic	46.5	52.1	-5.6	
White, non-Hispanic	15.5	25.0	-9.5	
Other	6.7	3.0	3.7	
Male (%)	50.7	47.2	3.5	0.342
Average age (years)	14.8	15.0	-0.3 *	0.000
Overage for grade[b] (%)	29.2	43.7	-14.5 *	0.000
Language other than English spoken at home (%)	46.6	41.3	5.3	0.145
Mother's education level (%)				0.294
Did not finish high school	19.6	24.5	-4.9	
High school diploma or GED certificate	26.2	30.7	-4.5	
Completed some postsecondary education	30.5	27.4	3.1	
Don't know	23.7	17.4	6.2	
GRADE reading comprehension[c]				
Average standard score	85.3	83.9	1.4 *	0.002
Corresponding grade equivalent	*5.0*	*4.7*		
Corresponding percentile	*15*	*13*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)				
English language arts (ELA)	2.1	1.6	0.4 *	0.000
Math	1.9	1.4	0.5 *	0.000
Social studies	2.0	1.6	0.4 *	0.000
Science	2.0	1.7	0.3 *	0.001
Credits earned in core subjects (as a percentage of credits attempted)	87.6	77.9	9.7 *	0.000
Attendance rate (%)	94.5	92.5	1.9 *	0.002
Free and reduced-price lunch (%)	68.1	41.5	26.6 *	0.000
Joint test of difference between program groups[e] ($\chi 2 = 397.4$)			*	0.000
Sample size[f]	2,563	199		

(continued)

Appendix Table E.7a (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The program year is the year in which the students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. Estimated differences between the school records sample and students without school records data are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "School Records Sample" are the observed means for students in the school records sample. The values in the "No School Records Data" column are the regression-adjusted means for students without school records data, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between students in the school records sample and students missing school records data. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference at baseline between students with or without school records data, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full study sample.

The Enhanced Reading Opportunities Study

Appendix Table E.7b

Characteristics of Students: Differences Between Students With and Without School Records in the Follow-Up Year, Reading Apprenticeship Schools

Characteristic	School Records Sample	No School Records Data	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.413
Hispanic	31.0	28.3	2.6	
Black, non-Hispanic	46.2	49.5	-3.3	
White, non-Hispanic	16.0	16.8	-0.8	
Other	6.7	5.3	1.4	
Male (%)	49.9	52.9	-3.0	0.208
Average age (years)	14.7	15.1	-0.4*	0.000
Overage for grade[b] (%)	26.3	46.2	-19.9*	0.000
Language other than English spoken at home (%)	46.4	45.4	1.0	0.717
Mother's education level (%)				0.294
Did not finish high school	19.2	23.2	-4.0	
High school diploma or GED certificate	26.9	24.9	2.0	
Completed some postsecondary education	30.1	31.2	-1.2	
Don't know	23.9	20.7	3.2	
GRADE reading comprehension[c]				
Average standard score	85.3	84.4	1.0*	0.001
Corresponding grade equivalent	*5.0*	*4.8*		
Corresponding percentile	*15*	*14*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)				
English language arts (ELA)	2.1	1.8	0.3*	0.000
Math	1.9	1.6	0.3*	0.000
Social studies	2.1	1.8	0.3*	0.002
Science	2.0	1.7	0.3*	0.002
Credits earned in core subjects (as a percentage of credits attempted)	88.3	81.0	7.3*	0.003
Attendance rate (%)	94.9	91.6	3.3*	0.000
Free and reduced-price lunch (%)	68.8	57.8	11.0*	0.018
Joint test of difference between program groups[e] ($\chi 2 = 403.5$)			*	0.000
Sample size[f]	2,212	550		

(continued)

Appendix Table E.7b (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The follow-up year is the year following students' enrollment in the ERO class; it corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2. Estimated differences between the school records sample and students without school records data are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "School Records Sample" are the observed means for students in the school records sample. The values in the "No School Records Data" column are the regression-adjusted means for students without school records data, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between students in the school records sample and students missing school records data. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference at baseline between students with or without school records data, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full study sample.

The Enhanced Reading Opportunities Study

Appendix Table E.8a

Characteristics of Students:
Differences Between Students With and Without School Records in the Program Year,
Xtreme Reading Schools

Characteristic	School Records Sample	No School Records Data	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)			*	0.003
Hispanic	32.5	22.7	9.8	
Black, non-Hispanic	46.3	50.5	-4.2	
White, non-Hispanic	15.2	20.5	-5.3	
Other	6.0	6.3	-0.3	
Male (%)	50.1	53.4	-3.3	0.325
Average age (years)	14.8	14.9	-0.2 *	0.000
Overage for grade[b] (%)	28.3	42.4	-14.1 *	0.000
Language other than English spoken at home (%)	48.6	44.2	4.4	0.168
Mother's education level (%)			*	0.000
Did not finish high school	18.4	20.2	-1.8	
High school diploma or GED certificate	25.3	28.7	-3.5	
Completed some postsecondary education	34.1	31.6	2.4	
Don't know	22.3	19.4	2.8	
GRADE reading comprehension[c]				
Average standard score	85.3	84.4	1.0 *	0.010
Corresponding grade equivalent	*5.0*	*4.8*		
Corresponding percentile	*15*	*14*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)				
English language arts (ELA)	2.0	1.8	0.2 *	0.004
Math	1.9	1.6	0.3 *	0.000
Social studies	2.1	1.8	0.3 *	0.000
Science	1.9	1.8	0.2 *	0.035
Credits earned in core subjects (as a percentage of credits attempted)	87.7	82.2	5.5 *	0.012
Attendance rate (%)	95.0	92.9	2.0 *	0.005
Free and reduced-price lunch (%)	68.5	47.6	20.9 *	0.000
Joint test of difference between program groups[e] ($\chi 2 = 408.5$)			*	0.000
Sample size[f]	2,587	246		

(continued)

Appendix Table E.8a (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort1 and the 2006-2007 school year for Cohort 2. Estimated differences between the school records sample and students without school records data are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "School Records Sample" are the observed means for students in the school records sample. The values in the "No School Records Data" column are the regression-adjusted means for students without school records data, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between students in the school records sample and students missing school records data. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference at baseline between students with or without school records data, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full study sample.

The Enhanced Reading Opportunities Study

Appendix Table E.8b

Characteristics of Students: Differences Between Students With and Without School Records in the Follow-Up Year, Xtreme Reading Schools

Characteristic	School Records Sample	No School Records Data	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.698
Hispanic	31.7	31.6	0.1	
Black, non-Hispanic	46.4	47.7	-1.4	
White, non-Hispanic	15.8	15.2	0.7	
Other	6.1	5.5	0.6	
Male (%)	50.1	51.6	-1.5	0.604
Average age (years)	14.7	15.0	-0.3*	0.000
Overage for grade[b] (%)	25.4	44.9	-19.6*	0.000
Language other than English spoken at home (%)	48.2	48.1	0.1	0.976
Mother's education level (%)			*	0.000
Did not finish high school	17.0	24.3	-7.2	
High school diploma or GED certificate	26.1	23.4	2.8	
Completed some postsecondary education	26.1	23.4	2.8	
Don't know	22.4	20.8	1.6	
GRADE reading comprehension[c]				
Average standard score	85.4	84.8	0.7*	0.008
Corresponding grade equivalent	*5.1*	*4.9*		
Corresponding percentile	*16*	*15*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)				
English language arts (ELA)	2.1	1.8	0.3*	0.000
Math	1.9	1.6	0.3*	0.000
Social studies	2.1	1.8	0.3*	0.001
Science	2.0	1.7	0.2*	0.001
Credits earned in core subjects (as a percentage of credits attempted)	88.9	80.9	7.9*	0.000
Attendance rate (%)	95.4	92.2	3.3*	0.000
Free and reduced-price lunch (%)	68.2	63.7	4.5	0.182
Joint test of difference between program groups[e] ($\chi 2 = 415.5$)			*	0.000
Sample size[f]	2,224	609		

(continued)

Appendix Table E.8b (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The follow-up year is the school lyear following student's enrollment in the ERO class; it corresponds to the 2006-2007 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. Estimated differences between the school records sample and students without school records data are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "School Records Sample" are the observed means for students in the school records sample. The values in the "No School Records Data" column are the regression-adjusted means for students without school records data, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between students in the school records sample and students missing school records data. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference at baseline between students with or without school records data, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full study sample.

The Enhanced Reading Opportunities Study

Appendix Table E.9

Characteristics of Students with State Test Data Relative to Students in the Full Study Sample

Characteristic	Full Study Sample	State Test Samples							
		Program Year				Follow-Up Year			
		ELA	Social Studies	Science	Math	ELA	Social Studies	Science	Math
Baseline student survey and testing[a]									
Race/ethnicity (%)		*	*	*	*	*	*	*	*
Hispanic	31.0	52.2	41.5	26.0	39.8	46.6	36.5	35.6	40.1
Black, non-Hispanic	46.8	38.9	46.6	52.7	46.0	30.5	36.7	39.9	39.1
White, non-Hispanic	16.0	6.1	6.4	15.3	9.2	16.5	18.9	18.1	14.5
Other	6.2	2.9	5.5	6.0	4.9	6.4	8.0	6.5	6.3
Overage for grade[b] (%)	29.9	25.0*	45.0*	29.8	27.9	20.4*	24.9*	22.2*	24.3*
Language other than English spoken at home (%)	47.1	60.4*	56.7	44.0	53.2*	57.2*	52.3	50.9	54.1*
Mother's education level (%)		*	*	*	*	*	*		*
Did not finish high school	19.2	22.2	22.0	16.8	19.6	20.0	18.0	18.3	19.4
High school diploma or GED	26.0	24.9	22.5	26.4	25.4	24.6	25.8	26.1	26.0
Some postsecondary education	32.1	28.2	28.7	34.6	32.0	29.5	33.2	31.5	29.6
Don't know	22.6	24.7	26.8	22.2	23.0	25.9	23.0	24.0	25.0
GRADE reading comprehension[c]									
Average standard score	85.2	85.8*	86.0*	85.6	85.9*	85.8*	85.7*	85.6*	85.5*
School records prior to program year[d]									
GPA in core subjects (out of 4.0)									
English language arts (ELA)	2.02	2.24*	1.91	2.14*	2.14*	2.17*	2.15*	2.19*	2.16*
Math	1.87	2.08*	1.77	1.94	2.04*	2.02*	2.06*	2.06*	2.03*
Social studies	2.04	2.30*	1.97	2.12	2.27*	2.22*	2.24*	2.26*	2.21*
Science	1.94	2.17*	1.85	2.07*	2.07*	2.06*	2.06*	2.13*	2.09*
Attendance rate (%)	94.6	95.6*	95.2	95.9*	95.4*	95.3*	95.9*	95.8*	95.7*
Free and reduced-price lunch (%)	67.1	77.3*	76.1*	67.0	72.2*	66.3	66.3	69.9	73.8*
Joint test of difference between students with and without state test data ($\chi 2$)[e]		2,471.6*	1,829.9*	2,115.7*	2,199.5*	1,829.2*	1,236.2*	945.9*	1,309.9*
Sample size[f]	5,595	2,244	952	2,348	2,668	2,408	2,237	2,661	2,537

(continued)

Appendix Table E.9 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2. The values in this table are the observed means for students in the study sample and the analysis samples (that is, the two-year longitudinal sample and the state test subsamples).

A two-tailed t-test was used to test for differences between students in each state test sample and students excluded from the sample, controlling for the blocking of random assignment by school and cohort. Statistical significance is indicated (*) when the p-value is less than or equal to 5 percent.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

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

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference at baseline between students in each state test sample and students excluded from the sample. These tests are based on the characteristics included in this table as well as the following variables: standardized state test scores in core subjects areas (reading, math, science, and social studies) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, credits earned in core subjects, age at random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full study sample and the state test samples.

The Enhanced Reading Opportunities Study

Appendix Table E.10

Characteristics of Students with State Test Data Relative to Students in the Full Study Sample, Reading Apprenticeship Schools

Characteristic	Full Study Sample	State Test Subsamples: Program Year ELA	Program Year Social Studies	Program Year Science	Program Year Math	Follow-Up Year ELA	Follow-Up Year Social Studies	Follow-Up Year Science	Follow-Up Year Math
Baseline student survey and testing[a]									
Race/ethnicity (%)		*	*	*	*	*	*	*	*
Hispanic	30.4	52.5	37.1	23.6	39.4	48.2	32.7	36.2	41.7
Black, non-Hispanic	46.9	36.7	51.0	55.6	43.5	25.4	39.8	38.2	39.7
White, non-Hispanic	16.2	7.3	5.8	14.2	12.0	19.5	19.1	18.5	12.3
Other	6.4	3.5	6.0	6.6	5.1	6.8	8.4	7.1	6.3
Overage for grade[b] (%)	30.2	25.5	48.5*	32.0	28.7	19.9*	24.2*	21.3*	24.4*
Language other than English spoken at home (%)	46.2	59.6*	53.9	42.6	52.4	57.1*	48.7	51.5	54.3
Mother's education level (%)		*	*	*	*	*	*	*	*
Did not finish high school	20.0	22.2	21.9	17.7	18.9	20.9	17.9	19.8	21.2
High school diploma or GED	26.5	25.3	21.2	26.9	26.8	24.5	26.1	26.5	24.6
Some postsecondary education	30.3	26.7	28.2	32.8	29.8	27.6	32.7	28.5	26.9
Don't know	23.2	25.8	28.7	22.6	24.5	27.0	23.3	25.2	27.3
GRADE reading comprehension[c]									
Average standard score	85.2	85.7	86.2*	85.2	85.9*	86.1*	85.9*	85.7*	85.6*
School records prior to program year[d]									
GPA in core subjects (out of 4.0)									
English language arts (ELA)	2.03	2.24*	1.90	2.16*	2.17*	2.21*	2.18*	2.21*	2.18*
Math	1.87	2.08*	1.76	1.96	2.08*	2.02*	2.10*	2.07*	2.04*
Social studies	2.01	2.23*	1.91	2.08	2.23*	2.20*	2.22*	2.25*	2.17*
Science	1.96	2.17*	1.81	2.11	2.12*	2.11*	2.09	2.16*	2.10*
Attendance rate (%)	94.3	95.6*	94.9	95.6*	95.3*	95.2*	95.6*	95.6*	95.6*
Free and reduced-price lunch (%)	66.9	77.6*	78.9*	68.3	71.8	65.9	64.4	69.3	75.6*
Joint test of difference between students with and without state test data ($\chi 2$)[e]		1,168.4*	872.5*	910.6*	1,006.6*	957.6*	653.0*	521.8*	729.1*
Sample size[f]	2,762	1,053	447	1,151	1,263	1,140	1,089	1,269	1,226

(continued)

Appendix Table E.10 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2. The values in this table are the observed means for students in the study sample and the analysis samples (that is, the two-year longitudinal sample and the state test subsamples).

A two-tailed t-test was used to test for differences between students in each state test sample and students excluded from the sample, controlling for the blocking of random assignment by school and cohort. Statistical significance is indicated (*) when the p-value is less than or equal to 5 percent.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

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

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference at baseline between students in each state test sample and students excluded from the sample. These tests are based on the characteristics included in this table as well as the following variables: standardized state test scores in core subjects areas (reading, math, science, and social studies) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, credits earned in core subjects, age at random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full study sample and state test samples.

The Enhanced Reading Opportunities Study

Appendix Table E.11

Characteristics of Students with State Test Data Relative to Students in the Full Study Sample, Xtreme Reading Schools

Characteristic	Full Study Sample	State Test Subsamples							
		Program Year				Follow-Up Year			
		ELA	Social Studies	Science	Math	ELA	Social Studies	Science	Math
Baseline student survey and testing[a]									
Race/ethnicity (%)		*	*	*	*	*	*	*	*
Hispanic	31.6	52.0	45.3	28.2	40.3	45.2	40.2	35.0	38.6
Black, non-Hispanic	46.8	40.8	42.8	50.0	48.3	35.0	33.7	41.5	38.6
White, non-Hispanic	15.7	5.0	6.9	16.5	6.6	13.8	18.6	17.7	16.6
Other	6.0	2.3	5.0	5.3	4.8	6.0	7.5	5.9	6.3
Overage for grade[b] (%)	29.5	24.6	41.8*	27.7	27.3	20.8*	25.4	23.0*	24.3*
Language other than English spoken at home (%)	48.1	61.0*	59.2	45.3	54.0*	57.3*	55.7	50.3	53.9
Mother's education level (%)		*	*	*	*	*			*
Did not finish high school	18.5	22.2	22.2	16.0	20.2	19.2	18.1	17.0	17.8
High school diploma or GED	25.6	24.5	23.6	25.9	24.2	24.7	25.5	25.8	27.3
Some postsecondary education	33.9	29.5	29.2	36.2	33.9	31.2	33.6	34.2	32.1
Don't know	22.0	23.8	25.1	21.8	21.8	24.9	22.8	23.0	22.8
GRADE reading comprehension[c]									
Average standard score	85.3	85.9*	85.7	85.9*	85.9*	85.5*	85.5	85.5	85.4
School records prior to program year[d]									
GPA in core subjects (out of 4.0)									
English language arts (ELA)	2.00	2.24*	1.92	2.11	2.11*	2.13*	2.11	2.17*	2.14*
Math	1.87	2.08*	1.79	1.92	2.00*	2.01*	2.02*	2.06*	2.01*
Social studies	2.08	2.36*	2.02	2.15	2.30*	2.23*	2.25*	2.27*	2.24*
Science	1.92	2.17*	1.88	2.04	2.03	2.00	2.03	2.10*	2.09*
Attendance rate (%)	94.8	95.7*	95.5	96.1*	95.5*	95.4*	96.2*	96.0*	95.9*
Free and reduced-price lunch (%)	67.3	77.1*	73.7	65.8	72.7*	66.7	68.2	70.5	72.0
Joint test of difference between students with and without state test data ($\chi 2$)[e]		1,379.9*	1,037.5*	1,308.3*	1,302.2*	943.0*	677.1*	518.1*	649.1*
Sample size[f]	2,833	1,191	505	1,197	1,405	1,268	1,148	1,392	1,311

(continued)

Appendix Table E.11 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2. The values in this table are the observed means for students in the study sample and the analysis samples (that is, the two-year longitudinal sample and the state test subsamples).

A two-tailed t-test was used to test for differences between students in each state test sample and students excluded from the sample, controlling for the blocking of random assignment by school and cohort. Statistical significance is indicated (*) when the p-value is less than or equal to 5 percent.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

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

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference at baseline between students in each state test sample and students excluded from the sample. These tests are based on the characteristics included in this table as well as the following variables: standardized state test scores in core subjects areas (reading, math, science, and social studies) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, credits earned in core subjects, age at random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full study sample and the state test samples.

For example, in the year prior to random assignment, students with state test data had a higher GPA in core subjects and a higher attendance rate, on average, than students without state test data. Because less than 50 percent of students in the full study sample have state test data in any given year, the impact findings related to state test scores presented in this report may not be generalizable to either the full study sample or to students without state test data.

Comparison of the ERO group and the non-ERO Group in the School Records Sample

In order to examine whether random assignment is preserved in the analysis samples, Appendix Tables E.12a and E.12b compare the baseline characteristics of ERO and non-ERO students in the school records sample (program year and follow-up year samples), for all schools in the study. Appendix Tables E.13a through E.14b present this comparison for each reading program separately. As shown in these tables:

- ERO and non-ERO students in the school records sample are not systematically different from each other in terms of their baseline characteristics.

There is a high degree of similarity between the ERO and the non-ERO group in the school records sample. The only exception is that among students with school records in the follow-up year (Table E.12b), the ERO group had a slightly higher GPA in core subjects areas in the prior school year than the non-ERO group (2.04 compared with 1.98, p-value for the difference = 0.018). However, an omnibus chi-squared test indicates that overall, the ERO and the non-ERO group are not systematically different in terms of their baseline characteristics, whether in the program year or the follow-up year. This suggests that the school records sample preserves the balance that was achieved with random assignment for the full study sample and that differences in high school outcomes between the two groups reflect the impact of the ERO programs rather than preexisting differences in students' background characteristics and/or prior achievement. This is also true for each of the groups of schools using the two supplemental literacy programs (RAAL and Xtreme Reading).

Appendix Table E.15 examines the comparability of ERO and non-ERO students in each of the state test samples. Specifically, the table presents omnibus tests of the joint difference in baseline characteristics between the two program groups, for each of the state test samples used in this report. As indicated in this table:

- For the most part, there are no systematic baseline differences between ERO and non-ERO students in the state test samples. The one exception is that there is a systematic difference in the baseline characteristics of ERO and non-ERO students in the social studies test sample in the program year.

The Enhanced Reading Opportunities Study

Appendix Table E.12a

Characteristics of Students with School Records in the Program Year

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.724
Hispanic	32.1	31.9	0.2	
Black, non-Hispanic	45.6	46.6	-1.0	
White, non-Hispanic	15.9	15.1	0.8	
Other	6.4	6.4	0.0	
Male (%)	49.5	51.6	-2.1	0.131
Average age (years)	14.8	14.8	0.0	0.793
Overage for grade[b] (%)	28.9	28.4	0.5	0.694
Language other than English spoken at home (%)	47.5	47.9	-0.5	0.698
Mother's education level (%)				0.647
Did not finish high school	19.2	19.0	0.3	
High school diploma or GED certificate	25.4	26.0	-0.5	
Completed some postsecondary education	31.7	33.0	-1.3	
Don't know	23.7	22.1	1.5	
GRADE reading comprehension[c]				
Average standard score	85.2	85.5	-0.3	0.055
Corresponding grade equivalent	*5.0*	*5.1*		
Corresponding percentile	*15*	*16*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	2.00	1.96	0.04	0.071
English language arts (ELA)	2.05	2.02	0.03	0.379
Math	1.92	1.86	0.06	0.061
Social studies	2.08	2.04	0.04	0.138
Science	1.98	1.92	0.06*	0.045
Credits earned in core subjects (as a percentage of credits attempted)	88.2	87.0	1.3	0.057
Attendance rate (%)	94.8	94.6	0.2	0.234
Free and reduced-price lunch (%)	68.6	67.9	0.7	0.534
Joint test of difference between program groups[e] ($\chi 2 = 42.7$)				0.239
Sample size[f]	2,937	2,213		

(continued)

Appendix Table E.12a (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of students with school records in the program year.

The Enhanced Reading Opportunities Study

Appendix Table E.12b

Characteristics of Students with School Records in the Follow-Up Year

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.671
Hispanic	31.7	31.1	0.6	
Black, non-Hispanic	45.4	46.5	-1.2	
White, non-Hispanic	16.5	15.8	0.7	
Other	6.4	6.6	-0.2	
Male (%)	49.4	50.7	-1.3	0.379
Average age (years)	14.7	14.7	0.0	0.982
Overage for grade[b] (%)	25.6	26.1	-0.5	0.700
Language other than English spoken at home (%)	47.2	47.5	-0.3	0.844
Mother's education level (%)				0.622
Did not finish high school	17.9	18.6	-0.7	
High school diploma or GED certificate	26.3	26.4	-0.1	
Completed some postsecondary education	31.7	32.9	-1.2	
Don't know	24.0	22.0	2.0	
GRADE reading comprehension[c]				
Average standard score	85.3	85.5	-0.3	0.113
Corresponding grade equivalent	*5.0*	*5.1*		
Corresponding percentile	*15*	*16*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	2.04	1.98	0.06*	0.018
English language arts (ELA)	2.08	2.04	0.04	0.218
Math	1.96	1.89	0.06*	0.049
Social studies	2.11	2.05	0.06	0.062
Science	2.03	1.94	0.09*	0.005
Credits earned in core subjects (as a percentage of credits attempted)	89.1	87.8	1.3	0.053
Attendance rate (%)	95.2	95.1	0.1	0.382
Free and reduced-price lunch (%)	68.9	67.7	1.2	0.337
Joint test of difference between program groups[e] ($\chi 2 = 39.4$)				0.365
Sample size[f]	2,542	1,894		

(continued)

Appendix Table E.12b (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The follow-up year is the school year following students' enrollment in the ERO class; it corresponds to the 2006-2007 school year for Cohort 1. The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of students with school records in the follow-up year.

The Enhanced Reading Opportunities Study

Appendix Table E.13a

Characteristics of Students with School Records in the Program Year, Reading Apprenticeship Schools

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.543
Hispanic	31.0	31.0	0.0	
Black, non-Hispanic	45.5	47.4	-1.9	
White, non-Hispanic	16.3	15.5	0.9	
Other	7.2	6.2	1.0	
Male (%)	49.7	51.9	-2.1	0.287
Average age (years)	14.8	14.8	0.0	0.737
Overage for grade[b] (%)	28.9	29.3	-0.4	0.822
Language other than English spoken at home (%)	45.4	47.5	-2.1	0.245
Mother's education level (%)				0.989
Did not finish high school	19.9	19.3	0.7	
High school diploma or GED certificate	26.0	26.4	-0.3	
Completed some postsecondary education	30.6	30.4	0.2	
Don't know	23.4	23.9	-0.5	
GRADE reading comprehension[c]				
Average standard score	85.3	85.3	0.0	0.964
Corresponding grade equivalent	*5.0*	*5.0*		
Corresponding percentile	*15*	*15*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	2.01	1.95	0.06	0.100
English language arts (ELA)	2.07	2.02	0.05	0.275
Math	1.93	1.86	0.07	0.122
Social studies	2.05	2.00	0.05	0.232
Science	2.01	1.95	0.06	0.155
Credits earned in core subjects (as a percentage of credits attempted)	88.2	86.7	1.6	0.100
Attendance rate (%)	94.5	94.4	0.1	0.575
Free and reduced-price lunch (%)	69.5	66.3	3.2*	0.049
Joint test of difference between program groups[e] ($\chi 2 = 27.1$)				0.883
Sample size[f]	1,468	1,095		

(continued)

Appendix Table E.13a (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2.

The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of students with school records in the program year.

The Enhanced Reading Opportunities Study

Appendix Table E.13b

Characteristics of Students with School Records in the Follow-Up Year, Reading Apprenticeship Schools

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.453
Hispanic	30.9	30.3	0.6	
Black, non-Hispanic	45.1	47.5	-2.4	
White, non-Hispanic	16.7	16.0	0.7	
Other	7.2	6.1	1.1	
Male (%)	49.5	50.1	-0.6	0.774
Average age (years)	14.7	14.7	0.0	0.984
Overage for grade[b] (%)	25.9	26.8	-0.8	0.637
Language other than English spoken at home (%)	45.1	47.6	-2.4	0.200
Mother's education level (%)				0.994
Did not finish high school	18.9	19.5	-0.6	
High school diploma or GED certificate	27.1	26.4	0.7	
Completed some postsecondary education	30.2	30.2	-0.1	
Don't know	23.8	23.9	0.0	
GRADE reading comprehension[c]				
Average standard score	85.4	85.4	0.0	0.983
Corresponding grade equivalent	*5.1*	*5.1*		
Corresponding percentile	*16*	*16*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	2.03	1.97	0.06	0.093
English language arts (ELA)	2.09	2.05	0.04	0.403
Math	1.96	1.89	0.07	0.157
Social studies	2.07	2.01	0.06	0.170
Science	2.04	1.97	0.08	0.092
Credits earned in core subjects (as a percentage of credits attempted)	88.8	87.5	1.3	0.183
Attendance rate (%)	94.9	94.9	0.1	0.764
Free and reduced-price lunch (%)	70.6	66.1	4.6*	0.009
Joint test of difference between program groups[e] ($\chi 2 = 32.8$)				0.667
Sample size[f]	1,285	927		

(continued)

Appendix Table E.13b (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The follow-up year is the school year following students' enrollment in the ERO class; it corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of students with school records in the first follow-up year.

The Enhanced Reading Opportunities Study

Appendix Table E.14a

Characteristics of Students with School Records in the Program Year, Xtreme Reading Schools

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.690
Hispanic	33.3	32.9	0.4	
Black, non-Hispanic	45.6	45.8	-0.2	
White, non-Hispanic	15.5	14.7	0.8	
Other	5.6	6.6	-1.0	
Male (%)	49.3	51.4	-2.1	0.285
Average age (years)	14.8	14.7	0.0	0.480
Overage for grade[b] (%)	28.9	27.6	1.3	0.438
Language other than English spoken at home (%)	49.5	48.4	1.1	0.532
Mother's education level (%)				0.208
Did not finish high school	18.5	18.6	-0.1	
High school diploma or GED certificate	24.9	25.6	-0.7	
Completed some postsecondary education	32.7	35.5	-2.7	
Don't know	23.9	20.3	3.5*	
GRADE reading comprehension[c]				
Average standard score	85.1	85.7	-0.6*	0.005
Corresponding grade equivalent	*5.0*	*5.1*		
Corresponding percentile	*15*	*16*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	1.99	1.96	0.03	0.363
English language arts (ELA)	2.02	2.02	0.01	0.887
Math	1.91	1.87	0.05	0.270
Social studies	2.12	2.08	0.04	0.368
Science	1.96	1.90	0.06	0.157
Credits earned in core subjects (as a percentage of credits attempted)	88.2	87.2	1.0	0.299
Attendance rate (%)	95.1	94.8	0.3	0.254
Free and reduced-price lunch (%)	67.7	69.5	-1.8	0.290
Joint test of difference between program groups[e] ($\chi 2 = 47.0$)				0.125
Sample size[f]	1,469	1,118		

(continued)

Appendix Table E.14a (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2.

The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of students with school records in the program year.

The Enhanced Reading Opportunities Study

Appendix Table E.14b

Characteristics of Students with School Records in the Follow-Up Year, Xtreme Reading Schools

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.502
Hispanic	32.5	31.9	0.7	
Black, non-Hispanic	45.6	45.5	0.1	
White, non-Hispanic	16.3	15.5	0.8	
Other	5.6	7.1	-1.5	
Male (%)	49.2	51.3	-2.0	0.341
Average age (years)	14.7	14.7	0.0	0.959
Overage for grade[b] (%)	25.4	25.5	-0.1	0.940
Language other than English spoken at home (%)	49.3	47.4	1.9	0.313
Mother's education level (%)				0.274
Did not finish high school	17.0	17.7	-0.7	
High school diploma or GED certificate	25.6	26.5	-0.9	
Completed some postsecondary education	33.3	35.6	-2.4	
Don't know	24.2	20.2	4.0*	
GRADE reading comprehension[c]				
Average standard score	85.2	85.7	-0.5*	0.022
Corresponding grade equivalent	*5.0*	*5.1*		
Corresponding percentile	*15*	*16*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	2.04	1.98	0.06	0.097
English language arts (ELA)	2.07	2.03	0.04	0.365
Math	1.96	1.90	0.06	0.172
Social studies	2.15	2.09	0.06	0.207
Science	2.01	1.90	0.10*	0.022
Credits earned in core subjects (as a percentage of credits attempted)	89.4	88.1	1.3	0.159
Attendance rate (%)	95.5	95.3	0.2	0.344
Free and reduced-price lunch (%)	67.0	69.2	-2.2	0.229
Joint test of difference between program groups[e] ($\chi 2 = 41.4$)				0.285
Sample size[f]	1,257	967		

(continued)

Appendix Table E.14b (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The follow-up year is the school year following students' enrollment in the ERO class; it corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in fall 2005 for Cohort 1 (beginning of ninth grade) and spring 2006 for Cohort 2 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the school year immediately prior to the program year (2004-2005 for Cohort 1 and 2005-2006 for Cohort 2), when most students were in eighth grade. For Cohort 2 students in one district, school records are from two years prior to the program year (2004-2005), when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of students with school records in the follow-up year.

The Enhanced Reading Opportunities Study

Appendix Table E.15

Tests for Systematic Differences in Baseline Characteristics Between the ERO and Non-ERO Groups, Students in the School Records Sample with State Test Scores

	All Schools		Reading Apprenticeship Schools		Xtreme Reading Schools		All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Sample	$\chi 2$	P-Value	$\chi 2$	P-Value	$\chi 2$	P-Value	Sample Size	Sample Size	Sample Size
All schools									
Program year									
English language arts (ELA)	47.7	0.112	28.0	0.857	39.8	0.305	2,244	1,053	1,191
Social studies	54.2*	0.034	34.6	0.580	49.2	0.057	952	447	505
Science	41.4	0.286	31.2	0.738	48.3	0.082	2,348	1,151	1,197
Math	38.5	0.400	19.4	0.993	32.5	0.681	2,668	1,263	1,405
Follow-up year									
English language arts (ELA)	27.0	0.885	21.9	0.977	36.1	0.512	2,408	1,140	1,268
Social studies	37.7	0.435	32.3	0.690	49.1	0.088	2,237	1,089	1,148
Science	36.3	0.500	32.6	0.673	33.2	0.604	2,661	1,269	1,392
Math	39.6	0.355	27.1	0.885	48.5	0.080	2,537	1,226	1,311

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

Chi-squared tests were used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the following characteristics: race/ethnicity, gender, whether a student is overage for grade, whether a language other than English is spoken at home, mother's education level, GRADE reading comprehension score at baseline, GPA in each core subject area in the school year prior to random assignment (ELA, math, social studies, and science), credits earned in core subjects areas in the school year prior to random assignment, whether a student received free or reduced-price lunch, standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, indicators of missing data for all relevant student characteristics, and random assignment strata. The statistical significance of these tests is indicated (*) when the p-value is less than or equal to 5 percent.

Baseline Characteristics of the GRADE Respondent Sample

Impacts on reading achievement — which were the focus of the previous two study reports — are estimated based on the sample of students who completed the GRADE reading assessment in the spring of the program year (ninth grade). Nonresponse and baseline analyses were conducted for this GRADE respondent sample. The patterns of findings for this sample are similar to those discussed above for the school records sample, namely:

- There are systematic baseline differences between students who did and did not complete the GRADE assessment.

Specifically, an omnibus test comparing these two groups of students indicates that there are systematic differences between them in terms of their background characteristics. Most notably, nonrespondents are more likely than respondents to have characteristics associated with doing more poorly in school (such as being overage for ninth grade and having lower baseline test scores).

- ERO and non-ERO students in the GRADE respondent sample are not systematically different from each other in terms of their baseline characteristics.

As noted earlier in this appendix, there is a statistically significant difference between the ERO group and the non-ERO group in terms of the percentage of students who completed the GRADE assessment. This raises a concern as to whether, among students who completed the GRADE, there is also a difference between the characteristics of ERO and non-ERO students. However, as discussed in prior reports, an omnibus test indicates that, overall, there is no systematic difference between the characteristics of ERO and non-ERO students in the GRADE respondent sample.

Detailed findings on these analyses can be found in the two previous study reports.[3]

[3]In both previously released ERO reports (Kemple et al., 2008; Corrin et al., 2008)), the baseline characteristic analysis for the GRADE respondent sample can be found in Tables 2.4 and Appendix Table B.2.

Appendix F

Technical Notes for Impact Findings

This appendix discusses various technical issues related to the estimation of program impacts in this report. The first section provides the statistical model used to estimate the impact of the Enhanced Reading Opportunities (ERO) programs on school-based outcomes and presents the standard errors and confidence interval for the impact estimates presented in the body of the report. The second section presents tables of impact estimates that are *not* adjusted for students' baseline characteristics. The third section presents p-values for the key impact estimates that are adjusted for multiple hypothesis testing. The fourth section presents graphs of the distribution of reading comprehension and course performance outcomes for the ERO group and the non-ERO group. The final section includes tables of the standard deviations used to calculate effect sizes in this report, as well as other standard deviations that may be useful for future research.

Statistical Model for Estimating Impacts

The impact of the ERO programs on student outcomes is estimated by fitting the following regression model to the analysis samples:

$$Y_{it} = \beta_t T_i + \sum_K \lambda_{kt} B_{ki} + \sum_S \delta_s X_{si} + \varepsilon_{it} \qquad (1)$$

where:

Y_{it} = Outcome of interest for student i in year t (program year or follow-up year).

T_i = Indicator of ERO group membership (treatment status). This indicator is equal to 1 if student i was assigned to the ERO program and zero otherwise.

B_{ki} = Random assignment block indicators (school-by-cohort), equal to 1 if student i is in random assignment block k and zero otherwise.[1] These blocks are included in the model to capture a central feature of the research design in which random assignment was conducted separately for each school in each implementation year. Controlling for random assignment blocks in the model also accounts for the clustering of student outcomes by school and cohort, because it explains all of the between-school and between-cohort variation in student outcomes.

[1]There are 68 random assignment block indicators (34 schools * 2 implementation years).

X_{si} = Set of S pre-random assignment characteristics and prior achievement outcomes for student i. These covariates reduce within-school variation in the outcome measure, thereby increasing the precision of the impact estimates.[2]

ε_{it} = A within-student error term.

Therefore:

β_t = The estimated impact of the ERO programs on outcome Y in follow-up year t.

For the new findings presented in this report, robust standard errors are used when testing whether the estimated impact of the programs is equal to zero.[3] Tables F.1 to F.3 present the standard error and confidence intervals for the impact estimates in the ERO study.

Unadjusted Impact Estimates

As explained in the first section of this appendix, the statistical model used to estimate impacts controls for several measures of students' baseline characteristics and prior achievement. In theory, it is not strictly necessary to control for these baseline characteristics, because random assignment should ensure that students in the ERO and non-ERO group are similar in expectation at baseline, thereby producing unbiased estimates of program impacts. However, by

---

[2]As explained in Chapter 2, the following covariates are included in the statistical model to adjust for random differences at baseline between the ERO and the non-ERO group: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by their school district in the year prior to ERO participation, a baseline measurement of the outcome variable in the school year prior to ERO participation, and three indicators of missing data (one of each of the two state test variables and one for the prior measurement of the outcome).

The decision about which covariates to include in the model was made prior to starting the impact analysis, based on the predictive/explanatory power of the covariates. The first two covariates (overage status and GRADE pretest scores) were used in the impact analyses for the first two ERO study reports, and are included again for this round of analysis because they are important predictors of high school outcomes. Scores on standardized math and reading tests are also important predictors of academic outcomes, but could not be included in prior reports' analyses because these data were not available at the time of analysis. Finally, the impact model controls for a "pretest" measure of the outcome variable of interest, because one of the best predictors of an outcome variable is the value of that variable at an earlier point in time.

[3]Huber-White standard errors are used to account for heteroskedasticity, which in this case may result from the ERO programs having an impact on the variance of the outcome (that is, the variance in Y may differ for the ERO and non-ERO group as a result of the program).

The Enhanced Reading Opportunities Study

Appendix Table F.1

Impacts with Confidence Intervals on Reading Achievement, GRADE Respondent Sample

Outcome	Estimated Impact (S.E.)	95% Confidence Interval	Estimated Impact Effect Size (S.E.)	95% Confidence Interval	P-Value for Estimated Impact
All schools					
Reading comprehension					
Average standard score	0.88*	0.33 - 1.44	0.09*	0.03 - 0.14	0.002
	(0.28)		(0.03)		
Reading Apprenticeship schools					
Reading comprehension					
Average standard score	1.23*	0.44 - 2.02	0.12*	0.04 - 0.20	0.002
	(0.40)		(0.04)		
Xtreme Reading schools					
Reading comprehension					
Average standard score	0.56	-0.23 - 1.34	0.05	-0.02 - 0.13	0.165
	(0.40)		(0.04)		

SOURCE: MDRC calculations from the Enhanced Reading Opportunities GRADE assessment, administered at the end of ninth grade (spring 2006 for Cohort 1 and spring 2007 for Cohort 2).

NOTES: The sample sizes used in the analyses are 4,584 students for all schools, 2,255 students for Reading Apprenticeship schools, and 2,329 students for Xtreme Reading schools.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline and whether a student was overage for grade at random assignment. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation for the non-ERO group in the program year (all schools).

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 standard error of the impact estimate is in parentheses (S.E).

Rounding may cause slight discrepancies in calculating sums and differences.

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

The Enhanced Reading Opportunities Study

Appendix Table F.2

Impacts with Confidence Intervals on Grade Point Average (GPA) and Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample

Outcome	Estimated Impact (S.E.)	95% Confidence Interval	Estimated Impact Effect Size (S.E.)	95% Confidence Interval	P-Value for Estimated Impact
All schools					
Program year					
GPA in core subject areas	0.06* (0.02)	0.02 - 0.10	0.07* (0.02)	0.03 - 0.11	0.002
Credits earned in core subject areas (%)	0.56* (0.24)	0.10 - 1.02	0.06* (0.02)	0.01 - 0.10	0.017
Follow-up year					
GPA in core subject areas	0.04 (0.02)	0.00 - 0.09	0.05 (0.02)	0.00 - 0.09	0.061
Cumulative credits earned in core subject areas (%)	0.53 (0.43)	-0.31 - 1.37	0.03 (0.03)	-0.02 - 0.08	0.212

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The sample sizes used in the analyses are 5,150 students in the program year and 4,436 students in the follow-up year.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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 standard error of the impact estimate is in parentheses (S.E.).

The sample sizes reported in this table are for students with school records data in the relevant year.

Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The cumulative number of credits earned is scaled as a percentage of the total number of core credits required for graduation in a student's district.

The Enhanced Reading Opportunities Study

Appendix Table F.3

Impacts with Confidence Intervals on Grade Point Average (GPA) and Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample, by Program

Outcome	Estimated Impact (S.E.)	95% Confidence Interval	Estimated Impact Effect Size (S.E.)	95% Confidence Interval	P-Value for Estimated Impact
Reading Apprenticeship schools					
Program year					
GPA in core subject areas	0.07* (0.03)	0.01 - 0.12	0.07* (0.03)	0.01 - 0.13	0.019
Credits earned in core subject areas (%)	0.61 (0.34)	-0.05 - 1.27	0.06 (0.03)	-0.01 - 0.13	0.072
Follow-up year					
GPA in core subject areas	0.03 (0.03)	-0.04 - 0.09	0.03 (0.04)	-0.04 - 0.10	0.392
Cumulative credits earned in core subject areas (%)	0.43 (0.62)	-0.78 - 1.64	0.03 (0.04)	-0.05 - 0.10	0.486
Xtreme Reading schools					
Program year					
GPA in core subject areas	0.06* (0.03)	0.00 - 0.12	0.06* (0.03)	0.00 - 0.12	0.038
Credits earned in core subject areas (%)	0.49 (0.33)	-0.16 - 1.13	0.05 (0.03)	-0.02 - 0.11	0.138
Follow-up year					
GPA in core subject areas	0.06 (0.03)	0.00 - 0.12	0.06 (0.03)	0.00 - 0.13	0.068
Cumulative credits earned in core subject areas (%)	0.66 (0.60)	-0.51 - 1.83	0.04 (0.03)	-0.03 - 0.11	0.266

(continued)

Appendix Table F.3 (continued)

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The sample sizes used in the analyses for Reading Apprenticeship schools are 2,563 students in the program year and 2,212 students in the follow-up year. The sample sizes used in the analyses for Xtreme Reading schools are 2,587 students in the program year and 2,224 students in the follow-up year.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent. The standard error of the impact estimate is in parentheses (S.E.).

The sample sizes reported in this table are for students with school records data in the relevant year.

Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The cumulative number of credits earned is scaled as a percentage of the total number of core credits required for graduation in a student's district.

including highly predictive student characteristics in the model, it is possible to improve the precision of the impact estimates.

In order to examine the extent to which estimated impacts on high school outcomes (presented in Chapter 4) are sensitive to the inclusion of these covariates, the statistical model was re-estimated *without* controlling for students' background characteristics and prior achievement.[4] Appendix Tables F.4 and F.5 present the "unadjusted" impact of the programs on grade point average (GPA) and credit accumulation in core subject areas, for all schools in the study and for each program separately (Reading Apprenticeship Academic Literacy and Xtreme Reading). These unadjusted estimates can be compared to the adjusted estimates in Tables F.2 and F.3. The results are as follows:

[4]These sensitivity tests still include random assignment blocks as fixed effects, in order to account for the way in which random assignment was conducted.

The Enhanced Reading Opportunities Study

Appendix Table F.4

Impacts on Grade Point Average (GPA) and Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), Not Adjusted for Student Baseline Characteristics, School Records Sample

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools					
Program year					
GPA in core subject areas	1.60	1.52	0.08*	0.09*	0.001
Credits earned in core subject areas (%)	21.4	20.8	0.6*	0.06*	0.015
Sample size	2,937	2,213			
Follow-up year					
GPA in core subject areas	1.59	1.51	0.07*	0.08*	0.007
Cumulative credits earned in core subject areas (%)	44.4	43.6	0.8	0.05	0.100
Sample size	2,542	1,894			

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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 sample sizes reported in this table are for students with school records data in the relevant year.

Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The cumulative number of credits earned is scaled as a percentage of the total number of core credits required for graduation in a student's district.

The Enhanced Reading Opportunities Study

Appendix Table F.5

Impacts on Grade Point Average (GPA) and Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), Not Adjusted for Student Baseline Characteristics, School Records Sample, by Program

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
Reading Apprenticeship schools					
Program year					
GPA in core subject areas	1.54	1.44	0.10*	0.11*	0.006
Credits earned in core subject areas (%)	20.5	19.8	0.8*	0.08*	0.039
Sample size	1,468	1,095			
Follow-up year					
GPA in core subject areas	1.55	1.48	0.06	0.07	0.105
Cumulative credits earned in core subject areas (%)	42.6	41.9	0.7	0.04	0.309
Sample size	1,285	927			
Xtreme Reading Schools					
Program year					
GPA in core subject areas	1.66	1.59	0.07	0.07	0.060
Credits earned in core subject areas (%)	22.3	21.8	0.5	0.05	0.172
Sample size	1,469	1,118			
Follow-up year					
GPA in core subject areas	1.62	1.54	0.09*	0.09*	0.027
Cumulative credits earned in core subject areas (%)	46.2	45.3	0.9	0.05	0.189
Sample size	1,257	967			

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The sample sizes reported in this table are for students with school records data in the relevant year.

Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The cumulative number of credits earned is scaled as a percentage of the total number of core credits required for graduation in a student's district.

- **Program year:** Like the adjusted estimates, the unadjusted impact estimates are statistically significant for both GPA in core subject areas (effect size = 0.09; p-value = 0.001) and for credits earned in core subject areas (effect size = 0.06; p-value = 0.015) in the program year.

- **Follow-up year:** Unlike the adjusted findings, the unadjusted impact estimates are statistically significant for GPA in core subject areas (effect size = 0.08; p-value = 0.007) in the follow-up year.

This difference between the adjusted and unadjusted estimates in the follow-up year stems from the fact that, among students with school records data in that year, ERO students had better prior academic performance (at baseline) than non-ERO students.[5] Because the impact model controls for students' baseline course performance, observed and unobserved differences in prior achievement between the ERO and non-ERO group are removed, thereby producing internally valid estimates of program impacts. Had student-level controls *not* been included in the statistical model, then the analysis would have produced an upwardly biased estimate of program impacts in the follow-up year. Thus, the student-level covariates included in the impact model not only add precision to the estimates, but they also protect the findings against bias resulting from baseline achievement differences between the ERO and non-ERO group.

Similar sensitivity tests for impacts on GRADE reading scores and reading behaviors were also conducted in the prior two study reports. As discussed in these reports, the magnitude of unadjusted impact estimates for reading outcomes was not appreciably different from that of adjusted estimates, though as expected the standard error of unadjusted estimates was larger. Detailed findings on these analyses can be found in Appendix E of prior reports.

Adjusted P-Values

As explained in Chapter 2, when impacts are estimated for several student outcomes, this increases the likelihood of concluding that a given impact estimate is statistically significant, when in fact the program has no impact on student outcomes (this is known as a "Type I" error or a "false positive"). In this report — which focuses on school-based outcomes — there are four primary indicators of program effectiveness: GPA in core subject areas in the program

[5]As shown in Table E.12b, there is a statistically significant difference in baseline GPA in core subject areas between ERO and non-ERO students for the follow-up year school records sample (difference = 0.06 point, p-value = 0.018). Although not statistically significant, the difference in core credits earned between the ERO and non-ERO groups for this sample of students is 1.3 credits (p-value = 0.053).

year and follow-up year, and credit accumulation in core subject areas in these two years. Hence, as a sensitivity test, p-values for estimated impacts on these four primary outcomes were adjusted for multiple hypothesis testing, to see whether such adjustments alter the conclusions made in Chapter 4.

These multiplicity adjustments were made using the resampling-based methods developed by Westfall and Young (1993) and outlined in Schochet (2008).[6] In the context of the ERO study, these methods are particularly appropriate, because they take into account the fact that the four primary outcomes are correlated to each other.[7] In contrast, other adjustment methods assume that outcomes (and therefore hypothesis tests) are independent, and therefore produce adjustments that are too conservative.

Appendix Table F.6 presents the adjusted p-values for the primary impacts in this report. As seen in this table:

- Adjusting the p-values for multiple hypothesis testing does not lead to different conclusions about whether key impacts on high school outcomes are statistically significant.

Distribution of Student Outcomes

As reported in Chapters 3 and 4, the ERO programs had a positive impact on students' reading comprehension scores and course performance outcomes. In addition to improving

---

[6]These methods have been used in other random assignment studies, for example, in the evaluation of the Moving to Opportunities (MTO) program (Kling, Liebman, and Katz, 2006).

[7]Specifically, resampling-based methods use bootstrapped samples to estimate the *joint* distribution of the primary outcomes under the null hypothesis (which in this case is that there is no impact). The p-values are then adjusted for multiplicity based on this joint distribution. In this report, adjusted p-values were obtained as follows:

1. The impact of the programs on each key outcome *k* was estimated based on the school records sample, and the residuals from these four regressions were added to the school records dataset.
2. A large number of datasets (bootstrap samples) were generated by randomly sampling students with replacement from the school records sample (10,000 bootstrapped samples).
3. For each of these samples, the impact of the ERO programs on the residual for each key outcome *k* was estimated, and the t-statistics for these estimates were stored.
4. The adjusted p-value for each primary outcome *k* was then calculated using Algorithm 3.3 in Westfall and Young (1993, pp. 85-86). In essence, this algorithm uses the t-statistics from the bootstrapped samples to approximate the joint distribution of all tests under the null hypothesis of no impact. The adjusted p-value for the estimated impact of the programs on outcome *k* is then obtained by examining the location of its t-statistic in this joint distribution.

The Enhanced Reading Opportunities Study

Appendix Table F.6

P-Values for Estimated Impacts on Primary Outcomes, Adjusted for Multiple Hypothesis Testing

Outcome	Estimated Impact	P-Value for Estimated Impact	Adjusted P-Value for Estimated Impact
Program year			
GPA in core subject areas	0.06	0.002	0.007
Credits earned in core subject areas (%)	0.6	0.017	0.047
Follow-up year			
GPA in core subject areas	0.04	0.061	0.113
Credits earned in core subject areas (%)	0.5	0.212	0.211

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation.

Adjusted p-values were obtained using the resampling methods described in Westfall and Young (1993, Algorithm 3.3, pp. 85-86), based on 10,000 replications with replacement.

The statistical significance of an impact estimate is indicated (*) when the p-value is less than or equal to 5 percent.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The cumulative number of credits earned is scaled as a percentage of the total number of core credits required for graduation in a student's district.

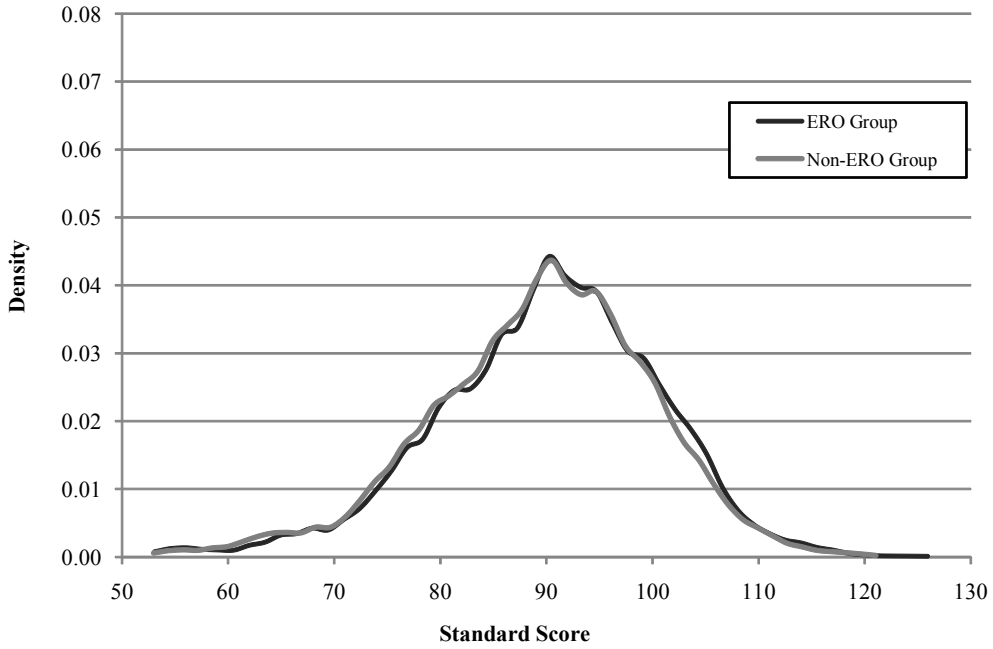
ERO students' *average* performance as a group, the ERO programs may have also had an impact on the *distribution* of ERO students' outcomes. This would occur, for example, if the ERO programs had a disproportionately larger impact for students at certain achievement levels.

To examine this issue, Appendix Figure F.1 plots the density of GRADE reading comprehension scores by program group. Figures F.2a to F.3b plot the density of GPA and credit

The Enhanced Reading Opportunities Study

Appendix Figure F.1

Density of GRADE Reading Comprehension Scores in the Program Year, by Program Group



SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2.

The density estimates in this figure are based on an Epanechnikov kernel with an optimal half-width, defined as $h=0.9*m/n^{1/5}$, where n is the sample size and m is the minimum of either (a) the standard deviation of the outcome or (b) the interquartile range of the outcome divided by 1.349. The sample sizes are 2,672 students for the ERO group and 1,912 students for the non-ERO group.

accumulation in core subject area, also by program group, and for each year (program year and follow-up year).[8] The dark line in these figures shows the distribution of outcomes for the ERO

[8]Density estimates are based on an Epanechnikov kernel with an optimal half-width, defined as:

$h = \frac{0.9m}{n^{1/5}}$ where

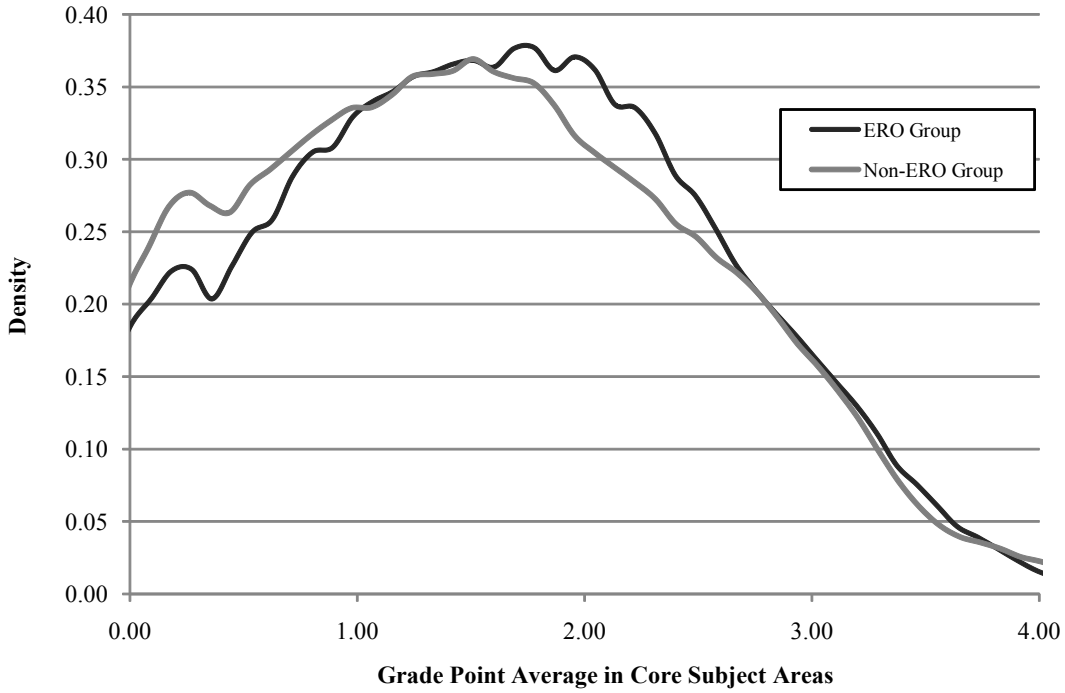
n = the relevant sample size and

m = the minimum of either (a) the standard deviation of the outcome of interest or (b) the interquartile range of the outcome of interest divided by 1.349.

The Enhanced Reading Opportunities Study

Appendix Figure F.2a

Density of Grade Point Average (GPA) in Core Subject Areas in the Program Year, by Program Group



SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The density estimates in this figure are based on an Epanechnikov kernel with an optimal half-width, defined as $h=0.9*m/n^{1/5}$, where n is the sample size and m is the minimum of either (a) the standard deviation of the outcome or (b) the interquartile range of the outcome divided by 1.349. The sample sizes are 2,937 students for the ERO group and 2,213 students for the non-ERO group.

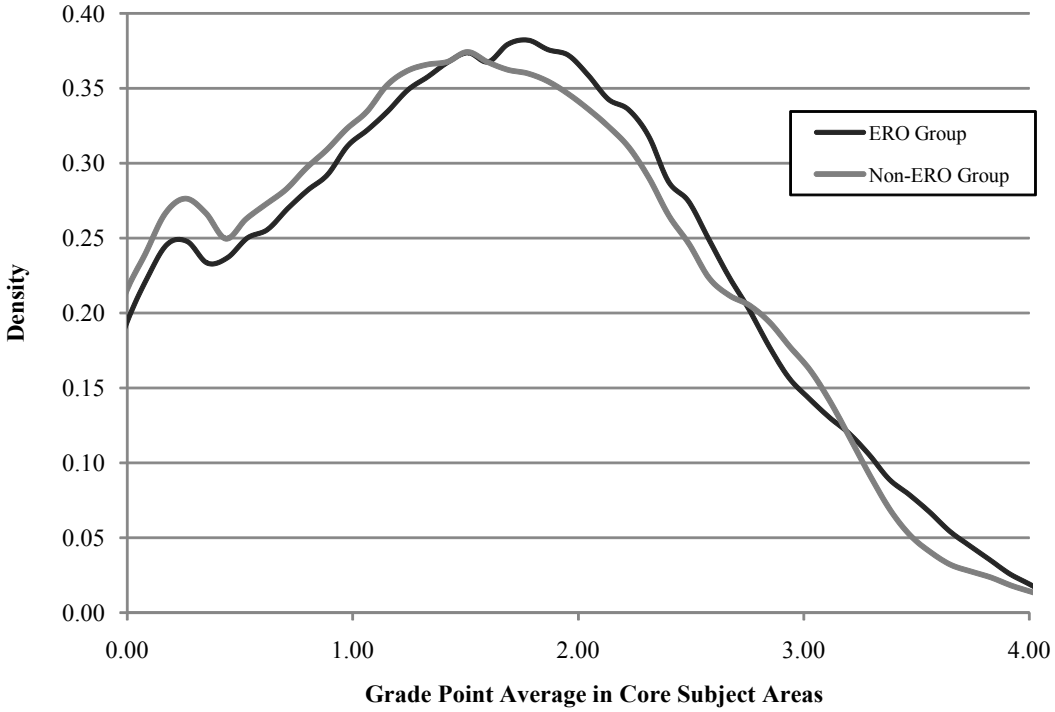
group, while the lighter line plots the distribution for the non-ERO group. The key findings from these figures are as follows:

- **Reading comprehension:** As seen in Figure F.1 (GRADE reading comprehension scores), the ERO curve is to the right of the non-ERO curve by an approximately equal amount at all points in the distribution of reading skills. This suggests that impacts on reading comprehension had a similar impact

The Enhanced Reading Opportunities Study

Appendix Figure F.2b

Density of Grade Point Average (GPA) in Core Subject Areas in the Follow-Up Year, by Program Group



SOURCE: MDRC calculations from school records data.

NOTES: The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The density estimates in this figure are based on an Epanechnikov kernel with an optimal half-width, defined as $h=0.9*m/n^{1/5}$, where n is the sample size and m is the minimum of either (a) the standard deviation of the outcome or (b) the interquartile range of the outcome divided by 1.349. The sample sizes are 2,542 students for the ERO group and 1,894 students for the non-ERO group.

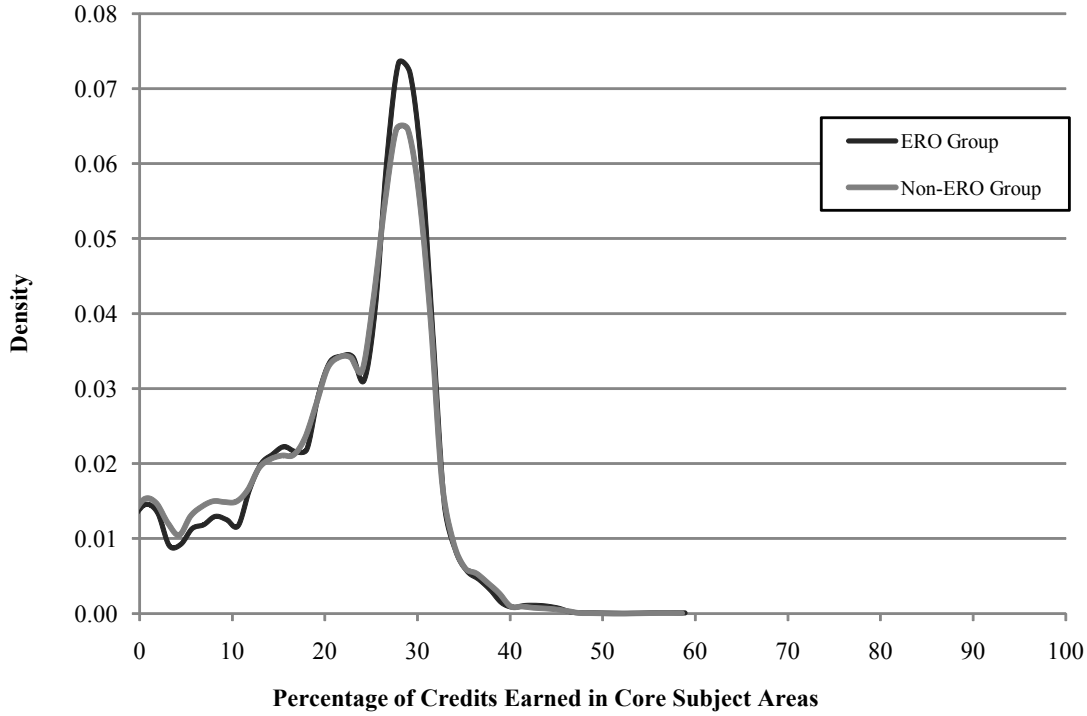
for all types of student, and that the programs did not affect the distribution of reading scores.

- **Course performance in the program year:** As seen in Figures F.2a and F.3a (GPA and credit accumulation in the program year), the ERO curve is indeed to the right of the non-ERO curve, *but only among students at the lower end of the course performance distribution*. As a result, the two groups' density curves have different shapes. This suggests that impacts on

The Enhanced Reading Opportunities Study

Appendix Figure F.3a

Density of Core Credit Accumulation in the Program Year, by Program Group



SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2.

The cumulative number of credits earned is scaled as a percentage of the total number of core credits required for graduation in a student's district.

The density estimates in this figure are based on an Epanechnikov kernel with an optimal half-width, defined as $h=0.9*m/n^{1/5}$, where n is the sample size and m is the minimum of either (a) the standard deviation of the outcome or (b) the interquartile range of the outcome divided by 1.349. The sample sizes are 2,937 for the ERO group and 2,213 for the non-ERO group.

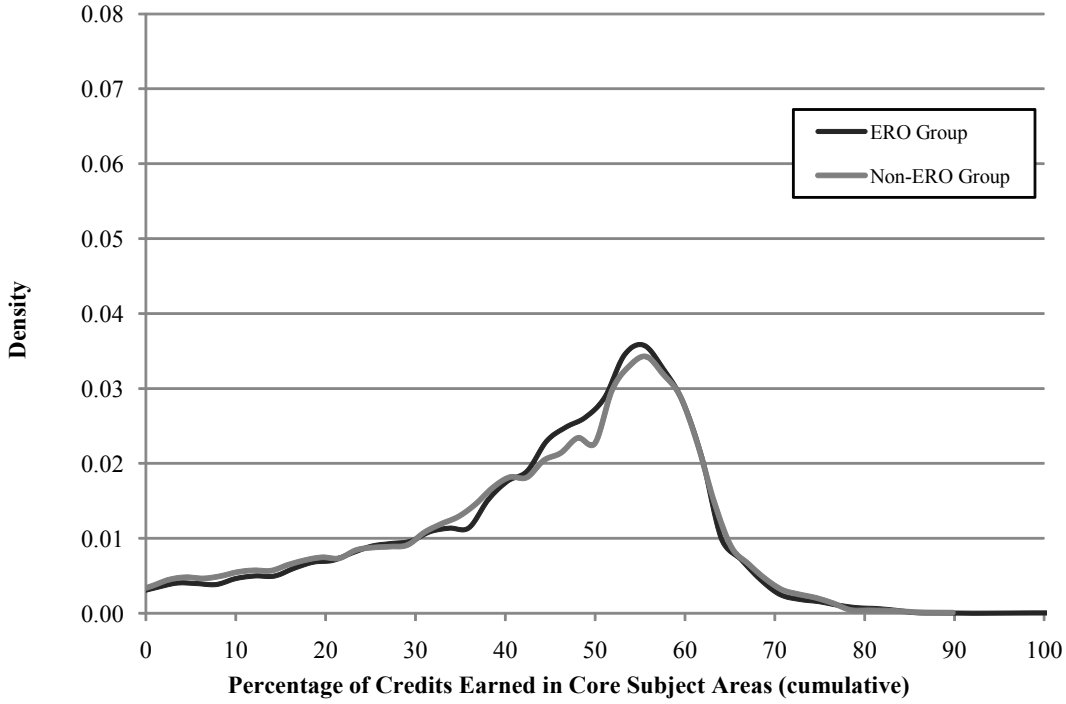
GPA and credit accumulation in the program year are mainly concentrated among students in the lower-achieving end of the course performance distribution. Hence, the ERO programs may have affected the distribution of course performance outcomes in the program year.

- **Course performance in the follow-up year (Figures F.2b and F.3b):** Impacts on course performance in the follow-up year also appear to be more

The Enhanced Reading Opportunities Study

Appendix Figure F.3b

Density of Core Credit Accumulation in the Follow-Up Year, by Program Group



SOURCE: MDRC calculations from school records data.

NOTES: The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The cumulative number of credits earned is scaled as a percentage of the total number of core credits required for graduation in a student's district.

The density estimates in this figure are based on an Epanechnikov kernel with an optimal half-width, defined as $h=0.9*m/n^{1/5}$, where n is the sample size and m is the minimum of either (a) the standard deviation of the outcome or (b) the interquartile range of the outcome divided by 1.349. The sample sizes are 2,542 for the ERO group and 1,894 for the non-ERO group.

concentrated among lower-achieving students, especially for GPA (Figure F.2b). However, because the estimated impact of the programs on GPA and credit accumulation is not statistically significant in the follow-up year, the extent of this distributional pattern is less pronounced in the follow-up year.

As explained in Appendix E, students excluded from the analysis due to a lack of data are lower performing on average than students included in the analysis sample. For this reason, Appendix E also warns that the impact findings in this report should not be generalized to

students *excluded from the analysis*. Figures F.1 and F.2 support this caution because they suggest that impacts for excluded students (who are lower performing) would have been larger than for students included in the analyses.

Standard Deviations Used to Calculate Effect Sizes

As explained in Chapter 2, impact estimates in this study are presented in both their original metric and as effect sizes. These effect sizes are based on the standard deviation of the outcome of interest for the non-ERO group in the relevant analysis sample.

Appendix Table F.7 presents the standard deviations used to calculate effect sizes in this report ("non-ERO" column).[9] In addition, the table also presents standard deviations for the ERO group — as well as for the ERO and non-ERO groups together — for use in future meta-analyses and research.[10]

[9]The standard deviations in this table are for both cohorts of participating ninth-grade students pooled together. These standard deviations are used to calculate effect sizes for the pooled impact of the ERO programs on reading achievement (Chapter 3) and on high school outcomes (Chapter 4).

Though similar, these are not the standard deviations used to calculate effect sizes for impacts on reading outcomes in the two previous reports. Effect sizes for reading outcomes in the first report are based on standard deviations for the non-ERO group in *Cohort 1*, while effect sizes in the second report are based on standard deviations for the non-ERO group in *Cohort 2*.

[10]Standard deviations for standardized state test scores are not presented because they are approximately equal to 1 by definition, due to the fact that these scores were standardized.

The Enhanced Reading Opportunities Study

Appendix Table F.7

Standard Deviations for the Non-ERO Group, ERO Group, and Pooled Groups, by Outcome and Year

Outcome	Program Year			Follow-Up Year			Second Follow-Up Year		
	Non-ERO Group	ERO Group	Pooled	Non-ERO Group	ERO Group	Pooled	Non-ERO Group	ERO Group	Pooled
GRADE reading assessment[a]									
Reading comprehension score	10.26	10.31	10.29	--	--	--	--	--	--
Reading vocabulary score	10.19	10.30	10.25	--	--	--	--	--	--
Reading behaviors									
Amount of school-related reading[b]	41.34	41.14	41.22	--	--	--	--	--	--
Amount of non-school-related reading[b]	32.46	31.20	31.73	--	--	--	--	--	--
Use of reading strategies[c]	0.64	0.62	0.63	--	--	--	--	--	--
GPA in core subject areas[d]	0.96	0.94	0.95	0.94	0.95	0.95	0.94	0.92	0.93
English language arts (ELA)	1.13	1.12	1.12	1.14	1.14	1.14	1.15	1.10	1.12
Social studies	1.17	1.15	1.16	1.17	1.16	1.17	1.14	1.11	1.12
Science	1.15	1.13	1.14	1.11	1.14	1.13	1.14	1.12	1.13
Math	1.13	1.12	1.12	1.13	1.13	1.13	1.12	1.12	1.12
Credits earned in core subject areas (%)[e]	9.87	9.62	9.73	17.04	16.39	16.67	20.77	20.21	20.44
English language arts (ELA)	10.90	10.56	10.71	17.69	16.65	17.11	21.46	19.82	20.51
Social studies	14.05	13.83	13.92	22.56	21.93	22.20	28.75	27.89	28.25
Science	15.63	15.15	15.36	25.79	25.73	25.76	30.96	32.38	31.80
Math	14.69	14.22	14.43	23.73	23.03	23.33	29.35	27.95	28.53
School behaviors									
Attendance rate (%)	9.91	9.49	9.67	14.50	14.12	14.29	16.20	14.91	15.45
Ever suspended (%)	47.26	46.65	46.92	47.62	46.73	47.12	45.57	45.34	45.42

(continued)

Appendix Table F.7 (continued)

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

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2. The second follow-up year corresponds to the 2007-2008 school year (available for Cohort 1 only).

Sample standard deviations in the table are for both cohorts of participating students pooled together (Cohort 1 and Cohort 2).

[a]The national average for standard scores is 100, and the standard deviation is 15.

[b]Frequency of reading is based on prior-month occurrences.

[c]Students' use of reading strategies is based on a 4-point scale.

[d]GPA is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

[e]Credit-earning variables are defined as the cumulative number of credits earned, scaled as a percentage of the total number of core credits required for graduation in a student's district.

Appendix G

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 Enhanced Reading Opportunities (ERO) programs. To assist in the planning of future studies on adolescent literacy, this appendix also presents estimates of the parameters that researchers need to know when conducting a statistical power analysis, which are based on the data collected in this study.

The discussion that follows is based on two related concepts that are often used to convey statistical power, that is, the "minimum detectable effect" (MDE) and the "minimum detectable effect size" (MDES). The MDE is the smallest true program impact that can be detected, given random sampling and estimation error.[1] The MDES is the minimum detectable effect scaled as an effect size — or in other words, it is the MDE divided by the standard deviation of the outcome of interest. Effect sizes are used widely for measuring the impacts of educational programs and are 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.

Calculating the Minimum Detectable Effect Size

The minimum detectable effect (MDE) and effect size (MDES) for a study are a function of the standard error of the estimated program impact:[2]

$$MDE = M_{N-B-X} * s.e.(\hat{\beta}) \tag{1a}$$

$$MDES = M_{N-B-X} * \frac{s.e.(\hat{\beta})}{\sigma} \tag{1b}$$

where:

[1] A 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.

[2] This is because the standard error of the impact estimate is what determines whether the impact estimate is statistically significant.

- $s.e.(\hat{\beta})$ is the standard error of the impact estimate
- σ is the standard deviation that is used to calculate effect sizes (for example, in this study, it is the standard deviation for the non-ERO group).
- N is the number of students in the sample
- B is the number of random assignment blocks in the impact analysis (in this study there are 34 schools * 2 cohorts = 68 blocks)
- X is the number of student baseline characteristics included as covariates in the impact model (of which there are eight in this report's analyses; see Appendix F)[3]
- M_{N-B-X} is the "degrees of freedom" multiplier, which is calculated to be 2.8 in this study, assuming a two-tailed test with a statistical power level 0.80 and a statistical significance level of 0.05.

The MDE and MDES presented in Chapter 2 of this report (Tables 2.6a to 2.6c) were calculated using these formulas, based on the standard error of the impact estimates, the sample size of the impact analyses, and the standard deviation of the non-ERO group for the relevant outcome (see Appendix Table F.4 for these standard deviations).

Planned Statistical Power in the Study Design Phase

In the planning phase of a study, the standard error of an impact estimate is not known so its value has to be approximated. Specifically, in the design phase, the MDES for a study can be estimated as follows:[4]

$$MDES = M_{N-J-X} * \sqrt{\frac{(1-\rho_s-\rho_d)(1-R_w^2)}{P(1-P)N} + \frac{\rho_s(1-R_s^2)}{P(1-P)J} + \frac{\rho_d(1-R_d^2)}{P(1-P)D}} \qquad (2)$$

[3]The following covariates are included in the impact analysis in this report to adjust for random differences at baseline between the ERO and the non-ERO group: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by their school district in the year prior to ERO participation, a baseline measurement of the outcome variable in the school year prior to ERO participation, and three indicators of missing data (one of each of the two state test variables and one for the prior measurement of the outcome).

[4]This minimum detectable effect size calculation assumes that the impact of the program is estimated using a fixed-effect specification (see previous appendix and Chapter 2 for details on this type of impact analysis). That is, the calculation assumes that the cross-site variation in the true impact of the program is zero.

where the parameters are defined as follows:

- P is the proportion of students randomly assigned to the treatment group
- N is the number of students in the sample
- J is the number of schools in the study
- D is the number of school districts in the study
- X is the number of student baseline characteristics to be included as covariates in the impact model
- ρ is the intraclass correlation:
  - ρ_d is the *district-level* intraclass correlation, defined as the proportion of the total variation in the outcome that is between districts.
  - ρ_s is the *school-level* intraclass correlation, defined as the proportion of the total variation in the outcome that is between schools within districts.
  - This means that $(1-\rho_s-\rho_d)$ is the proportion of total variation in the outcome that is within schools.
- R^2 is the explanatory power of the covariates in the impact model (including the treatment indicator and student baseline characteristics):
  - R_w^2 is the proportion of the *within-school* variance in the outcome that is explained by the covariates in the impact model
  - R_s^2 is the proportion of the *between-school* variance (within districts) that is explained by the covariates in the impact model
  - R_d^2 is the proportion of the *between-district* variance in the outcome that is explained by the covariates in the impact model
- M is the degrees of freedom multiplier, which takes a value of 2.8.

Notice that the three terms in Equation (2) represent three types of variation in the outcome measure. Specifically, the first component under the square root in Equation (2) accounts for variation in outcomes *within schools* (between students); the second component represents variation *between schools* within a district; and the third component measures variation *between districts*.

In the planning phase of the ERO study, the second and third components in Equation (2) were assumed to be equal to zero, because the evaluation uses a school fixed-effects regression model to estimate the impact of the programs (see Appendix F).[5] Thus, the calculation of the MDES simplified to the following equation in the planning phase:

$$MDES = 2.8 * \sqrt{\frac{(1-\rho_s-\rho_d)(1-R_w^2)}{P(1-P)(N)}} \qquad (3)$$

Based on informed assumptions about the values of the parameters in Equation (3),[6] the ERO study team chose a target sample size that would make it possible to achieve, for each cohort, an MDES of approximately 0.06 for all schools in the study and an MDES of approximately 0.10 for each of the ERO programs separately (for reading test scores as the outcome of interest).

It is important to note that the ERO study is not designed or powered to detect a *differential* impact between the two programs. Reading Apprenticeship Academic Literary (RAAL) and Xtreme Reading represent the same type of reading intervention and it was therefore expected that they would produce similar effects.[7] While tests of the difference in impact between the two programs are conducted, their purpose is mainly to verify that it is appropriate to estimate the impact of the two programs together as a class of intervention.

Intraclass Correlations and Variance Explained in this Study

To assist researchers with the planning of future studies on adolescent literacy, the parameter values in Equation (2) were estimated based on data collected for the ERO study:[8]

---

[5]This is because all of the variation in outcomes between schools (and by extension, between school districts) is explained by the statistical model

[6]The following parameter values were assumed for a reading test as the outcome measure, based on values from other studies and/or estimates provided by other researchers: $\rho_s + \rho_d = 0.07$; $P = 0.55$, and $R_w^2 = 0.69$. See Appendix C of the prior two study reports for details.

[7]Based on assumptions made in the study design phase, for example, the study was only powered to detect a differential impact of 0.30 on reading comprehension.

[8]These analyses are based on students in the school records sample who have data on the relevant outcome measure.

- Appendix Table G.1 presents the intraclass correlations for each outcome (ρ_s, ρ_d);[9]

- Appendix Table G.2 presents the proportion of variance explained by students' characteristics at baseline (R_w^2, R_s^2, R_s^2). The explanatory power of two sets of baseline characteristics is shown. This first is the explanatory power of the background characteristics and prior achievement included in the impact model ("all baseline measures").[10] The second is the explanatory power of the prior measurement of the outcome *by itself* ("prior measure of the outcome only").[11,12]

---

[9]Intraclass correlations were estimated by fitting the following random-effects model to the ERO data:

$$Y_{ijd} = \alpha_0 + \alpha_1 COH2 + \mu_d + \upsilon_{sd} + \varepsilon_{ijd}$$

where *Y* is the outcome of interest, *COH2* is an indicator for cohort, μ_d is a random error term for district *d* (i.e., the "between-district" residual), υ_{sd} is a random error term for school *j* in district *d* (i.e., the "between-school" residual), and ε_{ijd} is a random error term for student *i* in school *j* in district *d* (i.e., the "within-school" error term).

From this model, one can obtain a measure of variation in the outcome between districts (σ_μ^2), the variation between schools within districts (σ_υ^2), and the variation between students within schools (σ_ε^2).The following calculations were then performed:

District-level intraclass correlation (ρ_d): $\dfrac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\upsilon^2 + \sigma_\varepsilon^2}$

School-level intraclass correlation (ρ_s): $\dfrac{\sigma_\upsilon^2}{\sigma_\mu^2 + \sigma_\upsilon^2 + \sigma_\varepsilon^2}$

[10]See footnote 2 in Appendix F for a list of these characteristics.

[11]For example, for high school GPA as the outcome measure, this is the explanatory power of GPA in the 8th grade.

[12]The variance explained by these baseline characteristics was calculated by adding the relevant student characteristics to the model in footnote 9 and obtaining "adjusted" estimates of the variance between districts ($\tilde{\sigma}_\mu^2$), of the variation between schools within districts ($\tilde{\sigma}_\upsilon^2$), and of the variation between students within schools ($\tilde{\sigma}_\varepsilon^2$). The following calculations were then performed:

Within-school variance explained (R_w^2): $\dfrac{\sigma_\varepsilon^2 - \tilde{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}$

Between-school variance explained (R_s^2): $\dfrac{\sigma_\upsilon^2 - \tilde{\sigma}_\upsilon^2}{\sigma_\upsilon^2}$

Between-district variance explained (R_d^2): $\dfrac{\sigma_\mu^2 - \tilde{\sigma}_\mu^2}{\sigma_\mu^2}$

The Enhanced Reading Opportunities Study

Appendix Table G.1

School-Level and District-Level Intraclass Correlations for Student Outcomes

	Percentage of Variation:			
	Between Schools Within District (ρ_s)		Between Districts (ρ_d)	
Outcome	Program Year	Follow-Up Year	Program Year	Follow-Up Year
GRADE reading assessment[a]				
Reading comprehension score	0.04	NA	0.00	NA
Reading vocabulary score	0.01	NA	0.03	NA
Reading behaviors				
Amount of school-related reading[b]	0.02	NA	0.04	NA
Amount of non-school-related reading[b]	0.00	NA	0.02	NA
Use of reading strategies[c]	0.01	NA	0.04	NA
GPA in core subject areas[d]	0.06	0.04	0.04	0.04
English language arts (ELA)	0.10	0.05	0.00	0.02
Social studies	0.05	0.05	0.03	0.04
Science	0.05	0.03	0.05	0.03
Math	0.03	0.02	0.05	0.03
Credits earned in core subject areas (%)[e]	0.07	0.07	0.02	0.08
English language arts (ELA)	0.07	0.08	0.00	0.04
Social studies	0.10	0.05	0.04	0.16
Science	0.04	0.05	0.03	0.09
Math	0.02	0.04	0.04	0.05
State test score (standardized)[f]				
English language arts (ELA)	0.01	0.01	NA	NA
Social studies	0.02	0.03	NA	NA
Science	0.03	0.01	NA	NA
Math	0.04	0.02	NA	NA
School behaviors				
Attendance rate (%)	0.05	0.05	0.03	0.03
Ever suspended (%)	0.04	0.03	0.04	0.07

(continued)

Appendix Table G.1 (continued)

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

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The values in the first two columns of this table ("Between Schools Within District") are school-level intraclass correlations, or the proportion of the variation in the outcome measure that is between schools within district. The values in the last two columns ("Between Districts") are district-level intraclass correlations, or the proportion of the variation in the outcome measure that is between districts. To calculate these intraclass correlations, a random-effects model was used to decompose the variation in the outcome measure into three components: within schools, between schools, and between districts. These analyses are based on students in the school records sample who have data on the relevant outcome measure. See Appendix G for details.

There is no variation between districts in state test scores (NA), because these scores were standardized by school district.

[a]The national average for standard scores is 100, and the standard deviation is 15.

[b]Reading frequency is based on prior month occurrences.

[c]Students' use of reading strategies is based on a 4-point scale.

[d]GPA is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

[e]Credit-earning variables are defined as the cumulative number of credits earned, scaled as a percentage of the total number of core credits required for graduation in a student's district.

[f]State test scores are standardized by district, follow-up year, and cohort, using the mean and standard deviation of the non-ERO group in the relevant year.

As seen in these tables, the sum of the school-level and district-level intraclass correlation for the ERO study ($\rho_s + \rho_d$) is 0.04 for the GRADE reading comprehension test, which is similar to the value assumed in the planning phase of the study (0.07). The within-school explanatory power of students' characteristics (R_w^2) for the GRADE is 0.23 ("all baseline measures" column), which is lower than what was assumed in the planning phase (0.69).

The Enhanced Reading Opportunities Study

Appendix Table G.2

Variance Explained by Students' Background Characteristics and Prior Achievement

Outcome	Variance Explained Between Students Within Schools (R^2_w)		Variance Explained Between Schools Within District (R^2_s)		Variance Explained Between Districts (R^2_d)	
	All Baseline Measures	Prior Measure of Outcome Only	All Baseline Measures	Prior Measure of Outcome Only	All Baseline Measures	Prior Measure of Outcome Only
Program year						
GRADE reading assessment[a]						
Reading comprehension score	0.23	0.12	0.27	0.11	NA	NA
Reading vocabulary score	0.18	0.11	0.38	0.01	0.00	0.00
Reading behaviors						
Amount of school-related reading[b]	0.00	0.00	0.00	0.00	0.00	0.00
Amount of non-school-related reading[b]	0.00	0.00	0.17	0.02	0.02	0.06
Use of reading strategies[c]	0.00	0.00	0.01	0.01	0.00	0.00
GPA in core subject areas[d]	0.35	0.33	0.43	0.39	0.85	0.85
English language arts (ELA)	0.18	0.15	0.32	0.27	NA	NA
Social studies	0.20	0.17	0.22	0.18	0.86	0.91
Science	0.17	0.13	0.13	0.09	0.25	0.30
Math	0.20	0.16	0.07	0.00	0.48	0.56
Credits earned in core subject areas (%)[e]	0.19	0.13	0.28	0.09	0.00	0.00
English language arts (ELA)	0.08	0.04	0.11	0.00	NA	NA
Social studies	0.07	0.03	0.07	0.00	0.15	0.00
Science	0.08	0.03	0.17	0.06	0.00	0.00
Math	0.12	0.06	0.49	0.17	0.08	0.00
State test score (standardized)[f]						
English language arts (ELA)	0.24	0.20	0.44	0.43	NA	NA
Social studies	0.16	0.12	0.39	0.23	NA	NA
Science	0.13	0.09	0.10	0.00	NA	NA
Math	0.21	0.19	0.24	0.27	NA	NA

(continued)

Appendix Table G.2 (continued)

Outcome	Variance Explained Between Students Within Schools (R^2_w)		Variance Explained Between Schools Within District (R^2_s)		Variance Explained Between Districts (R^2_d)	
	All Baseline Measures	Prior Measure of Outcome Only	All Baseline Measures	Prior Measure of Outcome Only	All Baseline Measures	Prior Measure of Outcome Only
School behaviors						
Attendance rate (%)	0.31	0.28	0.57	0.53	0.36	0.43
Ever suspended (%)	0.11	0.09	-0.15	-0.07	0.72	0.54
Follow-up year						
GPA in core subject areas[d]	0.29	0.27	0.56	0.50	0.73	0.79
English language arts (ELA)	0.16	0.14	0.39	0.23	0.15	0.27
Social studies	0.17	0.15	0.26	0.20	0.57	0.60
Science	0.15	0.13	0.39	0.36	0.36	0.42
Math	0.15	0.12	0.15	0.03	0.58	0.66
Credits earned in core subject areas (%)[e]	0.17	0.13	0.26	0.11	0.00	0.00
English language arts (ELA)	0.07	0.04	0.16	0.00	0.07	0.00
Social studies	0.07	0.03	0.12	0.00	0.01	0.02
Science	0.07	0.03	0.20	0.06	0.01	0.11
Math	0.12	0.07	0.40	0.18	0.01	0.00
State test score (standardized)[f]						
English language arts (ELA)	0.19	0.13	0.51	0.35	NA	NA
Social studies	0.17	0.11	0.23	0.12	NA	NA
Science	0.15	0.07	0.24	0.01	NA	NA
Math	0.16	0.14	0.27	0.25	NA	NA
School behaviors						
Attendance rate (%)	0.14	0.12	0.48	0.37	0.00	0.00
Ever suspended (%)	0.09	0.08	-0.26	-0.21	0.59	0.53

(continued)

Appendix Table G.2 (continued)

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

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The values in the "All Baseline Measures" column represent the explanatory power of the baseline characteristics in the ERO impact analysis: the GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The values in the "Prior Measure of Outcome Only" column are the explanatory power of the baseline measurement of the outcome variable only.

To calculate the parameters in this table, a random-effects model was used to decompose the variation in the outcome measure into three components: within schools, between schools, and between districts. Student baseline characteristics were then added to the model, in order to examine the extent to which these three variance components decrease when the model controls for student characteristics. These analyses are based on students in the school records sample who have data on the relevant outcome measure. See Appendix G for details.

NA indicates that there is no variation between districts in the outcome to explain.

[a]The national average for standard scores is 100, and the standard deviation is 15.

[b]Reading frequency is based on prior-month occurrences.

[c]Students' use of reading strategies is based on a 4-point scale.

[d]GPA is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

[e]Credit-earning variables are defined as the cumulative number of credits earned, scaled as a percentage of the total number of core credits required for graduation in a student's district.

[f]State test scores are standardized by district, follow-up year, and cohort, using the mean and standard deviation of the non-ERO group in the relevant year.

Appendix H

Supplementary Impact Findings

This appendix provides additional impact findings not presented in the main body of the report. The first four tables display impact findings for secondary outcomes not included in the full report, while the remaining tables present estimated impacts on primary outcomes for different subsamples of students in the school records sample. The key findings from these tables are summarized below.[1]

Impacts on Additional Outcomes: Total Credits, Credits Attempted, and Proficiency on State Tests

As reported in Chapter 4, the Enhanced Reading Opportunities (ERO) programs had a positive impact on credit accumulation in core subject areas in the program year. However, the programs may also have had an impact on students' credit earning in elective courses. To investigate this question, Appendix Table H.1 presents the estimated impacts of the programs on students' total credit accumulation (all core and elective courses combined):

- There are positive impacts on students' total credit accumulation at the end of both the program year and the follow-up year. By the end of the follow-up year, ERO students had earned 1.0 percent more of the credits required for graduation than non-ERO students (effect size = 0.06; p-value = 0.017).

The ERO programs' impact on core credit accumulation could be due to differences in ERO and non-ERO students' course-taking patterns (as opposed to differences in their success at passing the courses that they do attempt). To examine this question, Appendix Tables H.2 and H.3 display estimated impacts on students' course-taking patterns (the credits that students attempted during each school year), for both the full sample of students and by program:

- There were no statistically significant differences between the ERO and non-ERO group in terms of the number of credits they attempted in combined core subject areas during the program year. However, non-ERO students did attempt 0.1 more core credits than ERO students by the end of the follow-up year (effect size = -0.05; p-value = 0.032).

This latter finding could be due in part to the fact that the ERO program had a positive impact on credits earned in the program year. That is, because non-ERO students earned fewer

[1]The p-values in these tables are not adjusted for multiple hypothesis testing, because they are not key outcomes in the evaluation (see Chapter 2 for a discussion of the approach to multiple hypothesis testing in this report).

The Enhanced Reading Opportunities Study

Appendix Table H.1

Impacts on Total Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample, by Program

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools					
End of program year					
Total credits earned (%)	26.1	25.2	0.9*	0.09*	0.000
Credits earned in core subject areas (%)	21.4	20.9	0.6*	0.06*	0.017
Credits earned in electives (%)	32.9	31.5	1.4*	0.10*	0.000
Sample size	2,937	2,213			
End of follow-up year (cumulative)					
Total credits earned (%)	53.6	52.6	1.0*	0.06*	0.017
Credits earned in core subject areas (%)	44.4	43.8	0.5	0.03	0.212
Credits earned in electives (%)	67.2	65.2	2.0*	0.08*	0.001
Sample size	2,542	1,894			
Reading Apprenticeship schools					
End of program year					
Total credits earned (%)	25.6	24.5	1.1*	0.12*	0.000
Credits earned in core subject areas (%)	20.5	19.9	0.6	0.06	0.072
Credits earned in electives (%)	33.3	31.2	2.1*	0.15*	0.000 †
Sample size	1,468	1,095			
End of follow-up year (cumulative)					
Total credits earned (%)	52.5	51.4	1.2	0.07	0.057
Credits earned in core subject areas (%)	42.6	42.1	0.4	0.03	0.486
Credits earned in electives (%)	67.3	64.7	2.6*	0.10*	0.004
Sample size	1,285	927			
Xtreme Reading schools					
End of program year					
Total credits earned (%)	26.6	26.0	0.6	0.06	0.065
Credits earned in core subject areas (%)	22.3	21.8	0.5	0.05	0.138
Credits earned in electives (%)	32.6	31.8	0.7	0.05	0.106
Sample size	1,469	1,118			
End of follow-up year (cumulative)					
Total credits earned (%)	54.8	53.9	0.9	0.05	0.134
Credits earned in core subject areas (%)	46.2	45.5	0.7	0.04	0.266
Credits earned in electives (%)	67.1	65.6	1.5	0.06	0.081
Sample size	1,257	967			

(continued)

Appendix Table H.1 (continued)

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The sample sizes reported in this table are for students with school records data in the relevant year.

The cumulative number of credits earned is scaled as a percentage of the total number of credits (total, core, or elective) required for graduation in a student's district.

The Enhanced Reading Opportunities Study

Appendix Table H.2

Credits Attempted in Core Subject Areas, School Records Sample

Outcome	ERO Group	Non-ERO Group	Estimated Difference	Estimated Difference Effect Size	P-Value for Estimated Difference
All schools					
End of program year					
Credits attempted in core subject areas	3.8	3.9	0.0	-0.03	0.262
English language arts (ELA)	1.0	1.0	0.0	0.03	0.246
Social studies	0.9	0.9	0.0	-0.03	0.274
Science	1.0	1.0	0.0*	-0.08*	0.002
Math	1.0	1.0	0.0	0.01	0.691
Sample size	2,937	2,213			
End of follow-up year (cumulative)					
Credits attempted in core subject areas	7.9	8.0	-0.1*	-0.05*	0.032
English language arts (ELA)	2.1	2.1	0.0	0.00	0.955
Social studies	1.9	1.9	0.0	-0.04	0.081
Science	1.9	2.0	0.0	-0.05	0.073
Math	2.0	2.0	0.0	-0.05	0.067
Sample size	2,542	1,894			

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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 sample sizes reported in this table are for students with school records data in the relevant year.

Rounding may cause slight discrepancies in calculating sums and differences.

The Enhanced Reading Opportunities Study

Appendix Table H.3

Credits Attempted in Core Subject Areas, School Records Sample, by Program

Outcome	ERO Group	Non-ERO Group	Estimated Difference	Estimated Difference Effect Size	P-Value for Estimated Difference
Reading Apprenticeship schools					
End of program year					
Credits attempted in core subject areas	3.8	3.8	0.0	-0.01	0.788
English language arts (ELA)	1.0	1.0	0.0	0.05	0.159
Social studies	0.9	0.9	0.0	-0.03	0.393
Science	0.9	1.0	0.0*	-0.09*	0.023
Math	1.0	1.0	0.0	0.05	0.187
Sample size	1,468	1,095			
End of follow-up year (cumulative)					
Credits attempted in core subject areas	7.8	7.9	-0.1	-0.05	0.161
English language arts (ELA)	2.0	2.0	0.0	0.01	0.815
Social studies	1.9	1.9	0.0	-0.06	0.082
Science	1.9	2.0	0.0	-0.06	0.093
Math	2.0	2.0	0.0	-0.01	0.828
Sample size	1,285	927			
Xtreme Reading schools					
End of program year					
Credits attempted in core subject areas	3.9	3.9	0.0	-0.05	0.177
English language arts (ELA)	1.0	1.0	0.0	0.01	0.859
Social studies	0.9	0.9	0.0	-0.02	0.496
Science	1.0	1.0	0.0*	-0.08*	0.041
Math	1.0	1.0	0.0	-0.03	0.385
Sample size	1,469	1,118			
End of follow-up year (cumulative)					
Credits attempted in core subject areas	8.0	8.1	-0.1	-0.06	0.104
English language arts (ELA)	2.1	2.1	0.0	-0.01	0.846
Social studies	1.9	1.9	0.0	-0.02	0.484
Science	2.0	2.0	0.0	-0.03	0.417
Math	2.0	2.1	0.0*	-0.10*	0.014
Sample size	1,257	967			

(continued)

Appendix Table H.3 (continued)

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The sample sizes reported in this table are for students with school records data in the relevant year.

Rounding may cause slight discrepancies in calculating sums and differences.

core course credits during the program year, they may have had to attempt more core courses in the follow-up year because they were further behind.

As reported in Chapter 4, the ERO programs had a positive impact on students' scores on English language arts (ELA) and math state tests in the program year. One question is whether the programs' impact on state test scores is such that they also positively affected whether students achieved "proficiency" on these tests. Accordingly, Appendix Table H.4 presents estimated program impacts on whether or not students met state proficiency standards in the two subject areas where the ERO programs had a positive impact on state test scores (that is, ELA and math scores during the program year):[2]

[2]For standards-based tests, proficiency is based on the cut-off used by the state for accountability purposes. For end-of course tests, "meeting proficiency" is defined as passing the test.

The Enhanced Reading Opportunities Study

Appendix Table H.4

Impacts on Meeting Proficiency Standards on State Tests,[a] School Records Sample, Program Year

Outcome[a]	Number of Students	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools						
English language arts (ELA) (%)	2,250	56.4	52.3	4.2	0.08*	0.020
Math (%)	2,672	34.8	32.6	2.2	0.05	0.141
Reading Apprenticeship schools						
English language arts (ELA) (%)	1,058	55.4	53.0	2.4	0.05	0.358
Math (%)	1,265	35.1	33.6	1.4	0.03	0.517
Xtreme Reading schools						
English language arts (ELA) (%)	1,192	57.3	51.4	5.9	0.12*	0.016
Math (%)	1,407	34.6	31.7	2.9	0.06	0.164

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, and a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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 numbers of students reported in this table are for students in the school records sample who have state test scores for a given subject area in the relevant year. If a student wrote a specific test more than once, only his or her first score is used.

Rounding may cause slight discrepancies in calculating sums and differences.

[a]The outcome is the percentage of students meeting or exceeding state proficiency standards. On end-of-course tests, students are considered proficient if they obtain a passing score as determined by the state department of education.

- The ERO programs had a positive impact on meeting state proficiency standards in ELA during the program year. This suggests that ERO programs not only had a positive impact on students' state test scores, but they also increased the percentage of students that successfully met state proficiency standards. The estimated impact on meeting state proficiency standards in math was not statistically significant.

It is important to note, however, that the impact on meeting proficiency in ELA appears to be driven by the fact that the "proficiency" cut-off on the ELA state test is set relatively low in Texas, which makes it more feasible for the ERO programs to have an impact on proficiency in this state.[3] Thus, the ERO programs' impact on meeting proficiency on ELA state tests may not generalize to states where proficiency standards are set such that struggling ninth-grade readers are further below the threshold for being proficient.

Impacts for Student Subgroups

As reported in Chapter 4, the ERO programs have a positive impact on course performance in the program year; the estimated impact in the follow-up year, though positive, is not statistically significant. One question that arises out of these findings is whether the lack of a statistically significant outcome in the follow-up year is due to the fact that the sample size decreases in the follow-up year because of attrition. To examine this issue, Appendix Tables H.5 through H.7 present program-year impacts on students' course performance (GPA and credits earned in core subject areas) and school behaviors, respectively, for the sample of students with school records data in the follow-up year (that is, students who have school records in both the program year and the follow-up year):

- Limiting the sample to students with school records data in both years does not change the pattern of findings in this report. For students with school records in the follow-up year, there are statistically significant impacts in the

[3]Specifically, 71 percent of Texas students in the analysis sample are proficient in ELA, while only 36 percent of students in other states are proficient in this subject area. This difference appears to be driven by a lower threshold for attaining ELA proficiency in Texas, rather than to a true difference in achievement between the Texas sample and the sample in other states. The mean GRADE score for students in Texas who achieved proficiency on the ELA state test is 91 standard score points, while the mean GRADE score for students who achieved proficiency on the ELA test in other states is 95 standard score points. This feature of the state test data affects the impact findings because Texas carries a relatively large weight in the analysis (53 percent of students who have an ELA test score in the program year are from Texas).

The Enhanced Reading Opportunities Study

Appendix Table H.5

Impacts on Grade Point Average (GPA) During the Program Year: Follow-Up-Year School Records Sample

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools					
GPA in core subject areas	1.66	1.60	0.06*	0.06*	0.004
English language arts (ELA)	1.78	1.70	0.08*	0.07*	0.010
Social studies	1.72	1.63	0.10*	0.08*	0.002
Science	1.60	1.52	0.08*	0.07*	0.007
Math	1.56	1.52	0.04	0.03	0.231
Sample size	2,542	1,894			
Reading Apprenticeship schools					
GPA in core subject areas	1.60	1.54	0.06*	0.06*	0.045
English language arts (ELA)	1.68	1.60	0.08	0.07	0.054
Social studies	1.71	1.61	0.10*	0.09*	0.019
Science	1.54	1.41	0.13*	0.11*	0.003
Math	1.48	1.49	-0.01	-0.01	0.833
Sample size	1,285	927			
Xtreme Reading schools					
GPA in core subject areas	1.73	1.66	0.06*	0.07*	0.038
English language arts (ELA)	1.87	1.80	0.07	0.06	0.083
Social studies	1.74	1.65	0.09*	0.07*	0.046
Science	1.65	1.62	0.04	0.03	0.376
Math	1.63	1.55	0.08*	0.07*	0.046
Sample size	1,257	967			

(continued)

Appendix Table H.5 (continued)

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The sample sizes reported in this table are for students with school records data in both the program year and the follow-up year. Because students may not have earned grades in all core subject areas in a given year, sample sizes differ for impacts in the specific core subject areas.

Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The Enhanced Reading Opportunities Study

Appendix Table H.6

Impacts on Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation) in the Program Year: Follow-Up-Year School Records Sample

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools					
Credits earned in core subject areas (%)	22.3	21.8	0.5*	0.05*	0.041
English language arts (ELA)	21.0	20.5	0.5	0.05	0.076
Social studies	21.8	21.3	0.5	0.03	0.194
Science	24.3	23.7	0.7	0.04	0.128
Math	23.2	22.7	0.6	0.04	0.156
Sample size	2,542	1,894			
Reading Apprenticeship schools					
Credits earned in core subject areas (%)	21.4	20.9	0.5	0.05	0.157
English language arts (ELA)	19.6	19.1	0.5	0.05	0.247
Social studies	21.6	21.3	0.2	0.02	0.651
Science	22.8	22.1	0.7	0.05	0.241
Math	22.7	21.9	0.8	0.06	0.151
Sample size	1,285	927			
Xtreme Reading schools					
Credits earned in core subject areas (%)	23.3	22.8	0.5	0.05	0.142
English language arts (ELA)	22.5	22.0	0.5	0.05	0.188
Social studies	22.0	21.2	0.7	0.05	0.145
Science	25.9	25.2	0.6	0.04	0.303
Math	23.8	23.5	0.3	0.02	0.564
Sample size	1,257	967			

(continued)

Appendix Table H.6 (continued)

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The sample sizes reported in this table are for students with school records data in both the program year and the follow-up year.

The cumulative number of credits earned is scaled as a percentage of the total number of credits (core or subject-specific) required for graduation in a student's district.

The Enhanced Reading Opportunities Study

Appendix Table H.7

Impacts on Student Behaviors During the Program Year: Follow-Up-Year School Records Sample

Outcome	Number of Students	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools						
Program year						
Attendance rate (%)	4,398	94.6	94.2	0.3	0.03	0.092
Ever suspended (%)	3,936	29.3	31.1	-1.7	-0.04	0.197
Reading Apprenticeship schools						
Program year						
Attendance rate (%)	2,186	94.1	93.6	0.5	0.05	0.077
Ever suspended (%)	2,016	28.4	29.9	-1.6	-0.03	0.394
Xtreme Reading schools						
Program year						
Attendance rate (%)	2,212	95.1	94.9	0.1	0.01	0.634
Ever suspended (%)	1,920	30.4	32.3	-1.9	-0.04	0.347

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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. Statistically significant differences in impacts between the two reading programs are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The numbers of students reported in this table are for students in the school records sample and with data on the specific student behavior outcomes in the relevant year. There are five schools with missing data for "ever suspended."

Rounding may cause slight discrepancies in calculating sums and differences.

The attendance rate is defined relative to the number of days that the student was enrolled in the district.

"Ever suspended" measures the percentage of students suspended (in school or out of school) or expelled one or more times during the relevant school year.

program year on grade point average (GPA) in core subject areas (effect size = 0.06; p-value = 0.004) and credits earned in core subject areas (effect size = 0.05; p-value = 0.041). There are no statistically significant impacts on student behaviors.

Finally, Appendix Tables H.8 through H.10 present estimated impacts for subgroups of students based on their reading comprehension score at baseline, whether a student is overage for grade, and whether a language other than English is spoken at home, respectively. For these groups of students, it cannot be concluded that program impacts were larger for one group than another for the following reasons:

- **Baseline reading comprehension.** There were positive impacts on both GPA[4] and credit accumulation[5] during the program year, for the subgroup of students who scored the lowest on the pretest. However, these impacts do not differ significantly from the estimated impacts on the other two groups of students, and therefore, it cannot be concluded that the impacts for any of these groups of students are different from the other two groups.

- **Overage for grade.** There were positive impacts on GPA in core subject areas for students who were overage for grade,[6] as well as students who were not overage for grade[7] during the program year. The difference in impacts between these two groups is not statistically significant, so it cannot be concluded that the impacts differed between students who were overage and not overage for grade.

- **Language spoken at home.** There were positive impacts on GPA[8] during both the program year and the follow-up year for students who come from multilingual families, but these impacts are not statistically different from the estimated impacts of the programs on GPA for students from English-only families (which are positive but not statistically significant). Therefore, it cannot be concluded that the effects of the ERO programs on these two groups are different.

---

[4]Effect size = 0.07 (p-value = 0.035).
[5]Effect size = 0.10 (p-value = 0.020).
[6]Effect size = 0.09 (p-value = 0.025).
[7]Effect size = 0.06 (p-value = 0.028).
[8]Effect size = 0.09 (p-value = 0.004) during the program year, and effect size = 0.08 (p-value = 0.040) during the follow-up year.

The Enhanced Reading Opportunities Study

Appendix Table H.8

Impacts on Grade Point Average (GPA) and Credit Accumulation in Core Subject Areas (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample, by Baseline Reading Comprehension Performance

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
2.0-3.0 years below grade level					
Program year					
GPA in core subject areas	1.80	1.75	0.05	0.05	0.206
Credits earned in core subject areas (%)	23.2	22.5	0.7	0.07	0.069
Sample size	954	759			
Follow-up year					
GPA in core subject areas	1.77	1.69	0.07	0.08	0.071
Cumulative credits earned in core subject areas (%)	47.4	46.4	1.0	0.06	0.162†
Sample size	850	658			
3.1-4.0 years below grade level					
Program year					
GPA in core subject areas	1.60	1.53	0.07	0.08	0.060
Credits earned in core subject areas (%)	21.2	21.3	-0.2	-0.02	0.722
Sample size	843	630			
Follow-up year					
GPA in core subject areas	1.59	1.53	0.06	0.06	0.187
Cumulative credits earned in core subject areas (%)	43.9	45.0	-1.1	-0.06	0.177
Sample size	728	534			
4.1-5.0 years below grade level					
Program year					
GPA in core subject areas	1.43	1.36	0.07*	0.07*	0.035
Credits earned in core subject areas (%)	20.1	19.2	0.9*	0.10*	0.020
Sample size	1,140	824			
Follow-up year					
GPA in core subject areas	1.42	1.41	0.01	0.01	0.838
Cumulative credits earned in core subject areas (%)	42.0	40.7	1.3	0.08	0.072
Sample size	964	702			

(continued)

Appendix Table H.8 (continued)

SOURCE: MDRC calculations from individual students' school records data supplied by each school district.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The first follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation , and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group in a given follow-up year in the full two-year longitudinal sample (all schools).

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. Statistically significant differences in impacts across subgroups are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The sample sizes reported in this table are for students with school records data in the relevant year. Because students may not have attempted credits in all core subject areas in a given year, sample sizes differ for impacts in the specific core subject areas.

Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The cumulative number of credits earned is scaled as a percentage of the total number of credits (core or subject-specific) required for graduation in a student's district.

The Enhanced Reading Opportunities Study

Appendix Table H.9

Impacts on Grade Point Average (GPA) and Credit Accumulation in Core Subject Areas (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample, by Whether Students Were Overage for Grade

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
Overage for grade					
Program year					
GPA in core subject areas	1.32	1.24	0.09*	0.09*	0.025
Credits earned in core subject areas (%)	18.9	18.5	0.4	0.04	0.403
Sample size	849	632			
Follow-up year					
GPA in core subject areas	1.31	1.25	0.06	0.07	0.193
Cumulative credits earned in core subject areas (%)	39.9	39.1	0.8	0.05	0.440
Sample size	652	493			
Not overage for grade					
Program year					
GPA in core subject areas	1.71	1.66	0.05*	0.06*	0.028
Credits earned in core subject areas (%)	22.5	22.0	0.5	0.05	0.064
Sample size	2,088	1,581			
Follow-up year					
GPA in core subject areas	1.68	1.64	0.04	0.04	0.155
Cumulative credits earned in core subject areas (%)	45.9	45.6	0.3	0.02	0.479
Sample size	1,890	1,401			

(continued)

Appendix Table H.9 (continued)

SOURCE: MDRC calculations from individual students' school records data supplied by each school district.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation , and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group in a given follow-up year in the full two-year longitudinal sample (all schools).

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. Statistically significant differences in impacts between the two subgroups are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The sample sizes reported in this table are for students with school records data in the relevant year. Because students may not have attempted credits in all core subject areas in a given year, sample sizes differ for impacts in the specific core subject areas.

Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The cumulative number of credits earned is scaled as a percentage of the total number of credits (core or subject-specific) required for graduation in a student's district.

The Enhanced Reading Opportunities Study

Appendix Table H.10

Impacts on Grade Point Average (GPA) and Credit Accumulation in Core Subject Areas (Credits Earned as a Percentage of Credits Required for Graduation), School Records Sample, by Language Spoken at Home

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
Students from multilingual families					
Program year					
GPA in core subject areas	1.68	1.59	0.09*	0.09*	0.004
Credits earned in core subject areas (%)	21.6	20.9	0.6	0.06	0.060
Sample size	1,394	1,056			
Follow-up year					
GPA in core subject areas	1.63	1.56	0.07*	0.08*	0.040
Cumulative credits earned in core subject areas (%)	44.1	42.9	1.2	0.07	0.054
Sample size	1,200	899			
Students from English-only families					
Program year					
GPA in core subject areas	1.53	1.48	0.05	0.05	0.080
Credits earned in core subject areas (%)	21.3	20.9	0.5	0.05	0.177
Sample size	1,543	1,157			
Follow-up year					
GPA in core subject areas	1.54	1.53	0.01	0.01	0.669
Cumulative credits earned in core subject areas (%)	44.6	44.8	-0.2	-0.01	0.734
Sample size	1,342	995			

(continued)

Appendix Table H.10 (continued)

SOURCE: MDRC calculations from individual students' school records data supplied by each school district.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation , and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group in a given follow-up year in the full two-year longitudinal sample (all schools).

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. Statistically significant differences in impacts between the two subgroups are also indicated (†) when the p-value for this difference is less than or equal to 5 percent.

The sample sizes reported in this table are for students with school records data in the relevant year. Because students may not have attempted credits in all core subject areas in a given year, sample sizes differ for impacts in the specific core subject areas.

Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The cumulative number of credits earned is scaled as a percentage of the total number of credits (core or subject-specific) required for graduation in a student's district.

Appendix I

Baseline and Impact Findings, by Cohort

This appendix presents baseline and impact findings for each of the two cohorts of students who participated in the Enhanced Reading Opportunities (ERO) study. Recall from Chapter 1 that the ERO programs were implemented for two school years, such that two cohorts of ninth-grade students enrolled in the programs. The first cohort of ninth-grade students enrolled in the ERO programs during the 2005-2006 school year (Cohort 1), while the second cohort of students participated in the programs during the 2006-2007 school year (Cohort 2).[1] Note that the results for Cohort 1 presented in this appendix include an additional follow-up year, because of this cohort's earlier participation in the ERO programs.

The first set of tables in this appendix compares the baseline characteristics and prior achievement of the ERO and non-ERO groups in the school records sample, for each of the two cohorts separately. Tables I.1 through I.3 examine the baseline characteristics of Cohort 1 for each school year (program year, first follow-up year, and second follow-up year), while Tables I.4 and I.5 present similar findings, but for students in Cohort 2 (program year and first follow-up year). These tables indicate that:

- Overall differences in background characteristics and prior achievement between the ERO and non-ERO groups are not statistically significant for either cohort. A joint test of the difference in baseline characteristics between the two program groups is not statistically significant for either cohort, for any year (program year or follow-up years).[2]

The second set of tables in this appendix presents estimated impacts on the two primary outcomes in this report (grade point average [GPA] and credit accumulation in core subject areas), for each cohort separately. Table I.6 presents findings for students in Cohort 1 (program year, first follow-up year, and second follow-up year), while Table I.7 presents estimated impacts for students in Cohort 2 (program year and first follow-up year). In general, the cohort-specific findings are similar to the overall results for the pooled sample:

- For each cohort of students, the ERO programs had a positive impact on GPA in core subject areas in the program year. The effect size is 0.06 for Cohort 1 (p-value = 0.043) and 0.08 for Cohort 2 (p-value = 0.012). The difference in impacts between the two cohorts is not statistically significant, so

[1] As noted in Chapter 2, it was decided to pool the two cohorts for the analyses in this report, because estimated impacts on the GRADE reading comprehension test did not differ by a statistically significant amount across cohorts (effect size = 0.09 for Cohort 1 and 0.08 for Cohort 2).

[2] The only notable difference between the two program groups is that in the follow-up year for Cohort 1, the ERO group has a statistically higher GPA in science than the non-ERO group.

it cannot be concluded that the impact of the ERO programs on GPA differs across cohorts.

- The ERO programs had a positive impact on credits earned in core subject areas in the program year, for students in Cohort 1 (effect size = 0.08; p-value = 0.018). Though positive, the estimated impact on credit accumulation is not statistically significant for students in Cohort 2. However, the difference in impacts between the two cohorts on this outcome is not statistically significant, so it cannot be concluded that impacts on credit accumulation differ across cohorts.

- In the follow-up year, estimated impacts on course performance were not statistically significant for either cohort. Nor were there statistically significant impacts on GPA and credit accumulation for Cohort 1 in the second follow-up year.

The Enhanced Reading Opportunities Study

Appendix Table I.1

Characteristics of Students with School Records in the Program Year, Cohort 1

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.888
Hispanic	32.8	32.7	0.1	
Black, non-Hispanic	44.9	45.9	-1.0	
White, non-Hispanic	16.2	15.4	0.8	
Other	6.1	6.0	0.1	
Male (%)	49.4	50.2	-0.8	0.694
Average age (years)	14.8	14.8	0.0	0.410
Overage for grade[b] (%)	30.2	27.5	2.7	0.101
Language other than English spoken at home (%)	46.4	46.0	0.4	0.816
Mother's education level (%)				0.870
Did not finish high school	19.2	20.4	-1.3	
High school diploma or GED certificate	27.0	26.2	0.9	
Completed some postsecondary education	31.5	32.4	-0.9	
Don't know	22.3	21.0	1.3	
GRADE reading comprehension[c]				
Average standard score	85.8	86.1	-0.3	0.197
Corresponding grade equivalent	*5.1*	*5.2*		
Corresponding percentile	*16*	*17*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	2.02	1.99	0.03	0.404
English language arts (ELA)	2.02	2.00	0.01	0.771
Math	1.99	1.95	0.04	0.371
Social studies	2.13	2.10	0.04	0.404
Science	2.00	1.95	0.06	0.186
Credits earned in core subjects (as a percentage of credits attempted)	88.0	87.1	0.9	0.332
Attendance rate (%)	94.4	94.2	0.2	0.367
Free and reduced-price lunch (%)	68.7	67.8	1.0	0.538
Joint test of difference between program groups[e] ($\chi 2 = 32.2$)				0.692
Sample size[f]	1,545	1,162		

(continued)

Appendix Table I.1 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1.

The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in fall 2005 (beginning of ninth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records are from the 2004-2005 school year, when most students were in eighth grade.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of Cohort 1 students with school records in the program year.

The Enhanced Reading Opportunities Study

Appendix Table I.2

Characteristics of Students with School Records in the Follow-Up Year, Cohort 1

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.881
Hispanic	33.0	32.6	0.4	
Black, non-Hispanic	44.1	45.0	-0.9	
White, non-Hispanic	16.9	16.1	0.8	
Other	6.0	6.3	-0.3	
Male (%)	48.9	49.8	-0.8	0.691
Average age (years)	14.7	14.7	0.0	0.445
Overage for grade[b] (%)	27.0	25.3	1.7	0.335
Language other than English spoken at home (%)	46.8	45.5	1.3	0.468
Mother's education level (%)				0.660
Did not finish high school	18.5	20.1	-1.6	
High school diploma or GED certificate	28.0	26.1	1.9	
Completed some postsecondary education	31.1	32.8	-1.7	
Don't know	22.4	21.0	1.4	
GRADE reading comprehension[c]				
Average standard score	85.9	86.2	-0.2	0.313
Corresponding grade equivalent	*5.1*	*5.2*		
Corresponding percentile	*16*	*17*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	2.06	2.01	0.05	0.141
English language arts (ELA)	2.05	2.03	0.02	0.608
Math	2.02	1.97	0.05	0.246
Social studies	2.17	2.12	0.05	0.238
Science	2.05	1.95	0.10*	0.025
Credits earned in core subjects (as a percentage of credits attempted)	89.2	88.3	1.0	0.322
Attendance rate (%)	94.9	94.8	0.1	0.811
Free and reduced-price lunch (%)	69.4	68.1	1.3	0.438
Joint test of difference between program groups[e] ($\chi 2 = 28.5$)				0.841
Sample size[f]	1,333	975		

(continued)

Appendix Table I.2 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The follow-up year is the school year following students' enrollment in an ERO class; it corresponds to the 2006-2007 school year for Cohort 1.

The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in fall 2005 (beginning of ninth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records are from the 2004-2005 school year, when most students were in eighth grade.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of Cohort 1 students with school records in the follow-up year.

The Enhanced Reading Opportunities Study

Appendix Table I.3

Characteristics of Students with School Records in the Second Follow-Up Year, Cohort 1

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.694
Hispanic	32.1	30.8	1.3	
Black, non-Hispanic	44.1	45.6	-1.5	
White, non-Hispanic	18.0	17.2	0.7	
Other	5.9	6.3	-0.4	
Male (%)	47.9	50.0	-2.1	0.364
Average age (years)	14.7	14.6	0.0	0.456
Overage for grade[b] (%)	22.4	20.8	1.6	0.390
Language other than English spoken at home (%)	46.7	43.8	2.9	0.160
Mother's education level (%)				0.324
Did not finish high school	16.8	18.5	-1.7	
High school diploma or GED certificate	29.7	26.0	3.7	
Completed some postsecondary education	31.5	34.9	-3.3	
Don't know	22.0	20.6	1.4	
GRADE reading comprehension[c]				
Average standard score	86.1	86.3	-0.3	0.281
Corresponding grade equivalent	*5.2*	*5.2*		
Corresponding percentile	*17*	*17*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	2.13	2.09	0.04	0.273
English language arts (ELA)	2.12	2.12	0.00	0.984
Math	2.09	2.03	0.06	0.223
Social studies	2.24	2.22	0.02	0.603
Science	2.11	2.02	0.09	0.064
Credits earned in core subjects (as a percentage of credits attempted)	90.6	90.0	0.7	0.504
Attendance rate (%)	95.5	95.4	0.1	0.578
Free and reduced-price lunch (%)	67.4	66.5	0.9	0.629
Joint test of difference between program groups[e] ($\chi 2 = 29.6$)				0.800
Sample size[f]	1,103	782		

(continued)

Appendix Table I.3 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The second follow-up year is the school year two years after students' enrollment in the ERO class; data for this follow-up year are only available for Cohort 1 (2007-2008 school year).

The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in fall 2005 (beginning of ninth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records are from the 2004-2005 school year, when most students were in eighth grade.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of Cohort 1 students with school records in the second follow-up year.

The Enhanced Reading Opportunities Study

Appendix Table I.4

Characteristics of Students with School Records in the Program Year, Cohort 2

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.857
Hispanic	31.4	31.0	0.4	
Black, non-Hispanic	46.3	47.4	-1.0	
White, non-Hispanic	15.7	14.8	0.9	
Other	6.6	6.8	-0.2	
Male (%)	49.6	53.2	-3.6	0.075
Average age (years)	14.7	14.8	0.0	0.593
Overage for grade[b] (%)	27.4	29.5	-2.0	0.246
Language other than English spoken at home (%)	48.6	50.1	-1.5	0.416
Mother's education level (%)				0.382
Did not finish high school	19.3	17.4	1.9	
High school diploma or GED certificate	23.8	25.7	-2.0	
Completed some postsecondary education	31.9	33.5	-1.7	
Don't know	25.1	23.3	1.8	
GRADE reading comprehension[c]				
Average standard score	84.5	84.8	-0.3	0.155
Corresponding grade equivalent	*4.9*	*4.9*		
Corresponding percentile	*14*	*15*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	1.98	1.93	0.06	0.081
English language arts (ELA)	2.08	2.04	0.04	0.337
Math	1.86	1.78	0.08	0.080
Social studies	2.03	1.98	0.05	0.204
Science	1.97	1.90	0.06	0.129
Credits earned in core subjects (as a percentage of credits attempted)	88.4	86.8	1.6	0.077
Attendance rate (%)	95.2	95.0	0.2	0.436
Free and reduced-price lunch (%)	68.5	68.0	0.5	0.788
Joint test of difference between program groups[e] ($\chi 2 = 45.9$)				0.151
Sample size[f]	1,392	1,051		

(continued)

Appendix Table I.4 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2006-2007 school year for Cohort 2.

The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in spring 2006 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the 2005-2006 school year, when most students were in eighth grade. For one district, school records are from the 2004-2005 school year, when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of Cohort 2 students with school records in the program year.

The Enhanced Reading Opportunities Study

Appendix Table I.5

Characteristics of Students with School Records in the Follow-Up Year, Cohort 2

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Baseline student survey and testing[a]				
Race/ethnicity (%)				0.804
Hispanic	30.3	29.4	0.8	
Black, non-Hispanic	46.7	48.2	-1.5	
White, non-Hispanic	16.1	15.5	0.7	
Other	6.9	6.9	-0.1	
Male (%)	49.9	51.8	-1.9	0.393
Average age (years)	14.7	14.7	0.0	0.380
Overage for grade[b] (%)	24.2	27.0	-2.8	0.123
Language other than English spoken at home (%)	47.6	49.6	-2.0	0.302
Mother's education level (%)				0.504
Did not finish high school	17.4	17.1	0.2	
High school diploma or GED certificate	24.6	26.7	-2.2	
Completed some postsecondary education	32.3	33.1	-0.7	
Don't know	25.7	23.1	2.7	
GRADE reading comprehension[c]				
Average standard score	84.5	84.8	-0.3	0.221
Corresponding grade equivalent	*4.9*	*4.9*		
Corresponding percentile	*14*	*15*		
School records prior to program year[d]				
GPA in core subjects (out of 4.0)	2.02	1.95	0.07	0.062
English language arts (ELA)	2.11	2.05	0.05	0.218
Math	1.89	1.82	0.08	0.106
Social studies	2.05	1.99	0.07	0.146
Science	2.00	1.92	0.08	0.080
Credits earned in core subjects (as a percentage of credits attempted)	88.9	87.3	1.7	0.075
Attendance rate (%)	95.6	95.3	0.2	0.275
Free and reduced-price lunch (%)	68.3	67.2	1.1	0.555
Joint test of difference between program groups[e] ($\chi 2 = 46.9$)				0.127
Sample size[f]	1,209	919		

(continued)

Appendix Table I.5 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline and school records data.

NOTES: The follow-up year is the school year following students' enrollment in the ERO class; it corresponds to the 2007-2008 school year for Cohort 2.

The estimated differences between the ERO group and the non-ERO group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort. The values in the column labeled "ERO Group" are the observed means for students randomly assigned to the ERO group. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed distribution of the ERO group across random assignment blocks as the basis for the adjustment.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. 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.

[a]Collected in spring 2006 (end of eighth grade).

[b]A student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

[c]The national average for standard scores 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). The italic type indicates that no statistical tests or arithmetic operations were performed on these reference points, because grade equivalents and percentiles are not equal-interval scales of measurement.

[d]School records for most students are from the 2005-2006 school year, when most students were in eighth grade. For one district, school records are from the 2004-2005 school year, when most of these students were in seventh grade, because these data were more consistently available.

[e]A chi-squared test was used to determine whether there is a systematic difference between the ERO group and the non-ERO group at baseline, based on the characteristics included in this table as well as the following variables: standardized state test scores in core subject areas (ELA, science, social studies, and math) in the school year prior to random assignment, the number of times a student was removed/expelled from school in the school year prior to random assignment, and indicators of missing data for all relevant student characteristics.

[f]Due to missing values, the number of students included varies by characteristic. The sample sizes reported here are for the full sample of Cohort 2 students with school records in the follow-up year.

The Enhanced Reading Opportunities Study

Appendix Table I.6

Impacts on Grade Point Average (GPA) and Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation) for Cohort 1, School Records Sample

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools					
Program year					
GPA in core subject areas	1.61	1.55	0.06*	0.06*	0.043
Credits earned in core subject areas (%)	21.7	20.9	0.8*	0.08*	0.018
Sample size	1,545	1,162			
Follow-up year					
GPA in core subject areas	1.57	1.51	0.05	0.06	0.094
Cumulative credits earned in core subject areas (%)	44.6	43.6	0.9	0.06	0.120
Sample size	1,333	975			
Second follow-up year					
GPA in core subject areas	1.80	1.76	0.04	0.04	0.276
Cumulative credits earned in core subject areas (%)	71.5	70.8	0.8	0.04	0.342
Sample size	1,103	782			

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year. The follow-up year corresponds to the 2006-2007 school year. The second follow-up year corresponds to the 2007-2008 school year.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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 sample sizes reported in this table are for students with school records data in the relevant year. Because students may not have earned grades in all core subject areas in a given year, sample sizes for GPA differ for impacts in the specific core subject areas.

Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The cumulative number of credits earned is scaled as a percentage of the total number of core credits required for graduation in a student's district.

The Enhanced Reading Opportunities Study

Appendix Table I.7

Impacts on Grade Point Average (GPA) and Credit Accumulation (Credits Earned as a Percentage of Credits Required for Graduation) for Cohort 2, School Records Sample

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
All schools					
Program year					
GPA in core subject areas	1.59	1.52	0.07*	0.08*	0.012
Credits earned in core subject areas (%)	21.2	20.9	0.3	0.03	0.357
Sample size	1,392	1,051			
Follow-up year					
GPA in core subject areas	1.61	1.57	0.03	0.04	0.320
Cumulative credits earned in core subject areas (%)	44.1	44.0	0.1	0.00	0.889
Sample size	1,209	919			

SOURCE: MDRC calculations from school records data.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2006-2007school year. The follow-up year corresponds to the 2007-2008 school year.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and cohort, as well as random baseline differences between the ERO and non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The "Non-ERO Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-ERO group, using the observed mean covariate values for the ERO group as the basis for the adjustment. The estimated impact effect size is calculated as a proportion of the standard deviation of the outcome for the non-ERO group during the relevant year (all schools).

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 sample sizes reported in this table are for students with school records data in the relevant year. Because students may not have earned grades in all core subject areas in a given year, sample sizes for GPA differ for impacts in the specific core subject areas.

Rounding may cause slight discrepancies in calculating sums and differences.

GPA in core subject areas is based on a 4-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

The cumulative number of credits earned is scaled as a percentage of the total number of core credits required for graduation in a student's district.

Appendix J

The Association Between Reading Outcomes and Academic Performance in High School

This appendix presents findings on the association between students' ninth-grade reading outcomes (reading achievement and reading behaviors) and their academic performance outcomes in ninth and tenth grade. The first section of this appendix describes the statistical model and the sample of students used to estimate these associations. The second section presents the findings.

Analytical Approach

The regression-adjusted association between students' academic performance and their ninth-grade reading outcomes is estimated by fitting the following model:

$$Y_{it} = \sum_{R} \alpha_{rt} READ_{ri} + \sum_{K} \lambda_{kt} B_{ki} + \sum_{S} \delta_{s} X_{si} + \varepsilon_{it} \quad (1)$$

where:

Y_{it} = Academic performance outcome of interest (that is, GPA, credits earned, state test scores) for student i in year t (program year or follow-up year).

$READ_{ri}$ = Set of five reading outcomes for student i at the end of the program year (that is, GRADE reading comprehension score, GRADE vocabulary scores, and three reading behaviors).[1]

B_{ki} = Random assignment block indicators (school-by-cohort), equal to 1 if student i is in random assignment block k and zero otherwise.[2] These blocks account for all observed and unobserved characteristics of schools and cohorts in the study.

X_{si} = Set of S pre-random assignment characteristics and prior achievement outcomes for student i. These are the same characteristics that are in-

[1]In addition to students' GRADE reading comprehension and vocabulary scores, the ERO study collected information (and estimated impacts) on three reading behaviors. These reading behaviors come from the student survey administered at the end of ninth grade; these are the frequency with which students read, both inside and outside of school, and their self-reported use of the reading strategies taught by the ERO programs. Reading frequency is measured as prior-month occurrences, while students' use of reading strategies is based on a 4-point scale. (See Appendix B for details on the scale and construction of these measures.)

[2]There are 68 random assignment block indicators (34 schools * 2 implementation years).

cluded in the impact model (see Appendix F), with the exception of the GRADE baseline score which is excluded from Model 1.[3]

ε_{it} = A within-student error term.

Therefore:

α_{rt} = The estimated association between reading outcome *R* and academic outcome *Y* in follow-up year *t*, *controlling for all other reading outcomes.* (Specifically, this coefficient represents the estimated effect of a 1-point increase in reading outcome *R* on outcome *Y*).

In order to compare associations across academic outcomes and years, these associations (α_{rt}) are converted to standardized regression coefficients (β_{rt}) as follows:

$$\beta_t = \frac{\alpha_{rt}}{\sigma_{Y,t}} * \sigma_{READ,t}$$

where:

$\sigma_{Y,t}$ = Standard deviation of outcome *Y* in year *t* among students in the analysis sample.

σ_{Rt} = Standard deviation of reading comprehension outcome *R* in year *t* among students in the analysis sample.

Therefore:

β_{rt} = The standardized association between reading outcome *R* and outcome *Y*. β_{rt} represents the effect of a 1 standard deviation increase in reading outcome *R* on outcome *Y*, scaled as an effect size (that is, as a proportion of the standard deviation in *Y*), controlling for all other reading outcomes.

Associations between students' reading and academic performance outcomes are based on the subset of students in the school records sample who have data on all five reading outcomes (the two reading achievement subtests and the three reading behaviors measures). This sample includes 4,293 students in the program year (representing 83 percent of students in the

[3]The GRADE reading comprehension baseline test score is not included because students' GRADE score at the end of ninth grade is already included as a covariate in the model.

school records sample in the program year) and 3,891 students in the follow-up year (representing 88 percent of students in the school records sample in the follow-up year).[4]

Associations Between Academic Performance and Reading Outcomes

Regression-adjusted associations are reported in Appendix Tables J.1 to J.3. These tables present the association between the five ninth-grade reading outcomes and students' grade point average [GPA] (J.1), credit accumulation (J.2), and state test scores (J.3). The values in these tables are standardized regression coefficients; that is, they represent the effect of a 1 standard deviation increase in the given reading outcome on the academic performance outcome, scaled as an effect size, and controlling for all other reading outcomes. Some of the general findings to note about these tables are:

- **Course performance:** Of the reading measures listed in Tables J.1 and J.2, GRADE reading comprehension scores are most strongly associated with course performance in terms of their magnitude (coefficient = 0.15, p-value = 0.00 for GPA and coefficient = 0.12, p-value = 0.000 for credit accumulation). Students' use of reflective reading strategies, as well as the amount of school-related reading they do, are also positively associated with course performance, though the magnitude of these associations are smaller (coefficients ranging from 0.02 – 0.09 for both GPA and credit accumulation). In contrast, reading vocabulary scores and the amount of non-school-related reading are in some cases negatively associated with students' course performance and the association is not statistically significant.

- **Performance on state tests:** The pattern of findings for state tests in Table J.3 is similar to that described above for course performance, with one exception: The direction of the association between state test scores and vocabulary scores is consistently positive.

- **Associations by follow-up year:** The association between ninth-grade reading comprehension scores and academic performance in the *follow-up year* (tenth grade) is smaller in magnitude than the association between reading scores and academic performance in the *program year* (ninth grade), but still

[4]The school records sample includes 5,150 students in the program year and 4,436 students in the follow-up year.

The Enhanced Reading Opportunities Study

Appendix Table J.1

Association Between Reading Outcomes and Grade Point Average in Core Subject Areas (Standardized Regression Coefficients)

Reading Predictors (End of Program Year)[a]	Core GPA in the Program Year: All Core	ELA	Social Studies	Science	Math	Core GPA in the Follow-Up Year: All Core	ELA	Social Studies	Science	Math
Reading achievement										
Reading comprehension										
Standardized coefficient	0.15*	0.13 *	0.14 *	0.14 *	0.13 *	0.11 *	0.10 *	0.15 *	0.10 *	0.07*
Standard error	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Reading vocabulary										
Standardized coefficient	-0.01	0.00	-0.01	-0.02	-0.03 *	0.03	0.01	0.01	-0.01	0.02
Standard error	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
P-value	0.616	0.913	0.589	0.161	0.042	0.105	0.736	0.542	0.545	0.204
Reading behaviors										
Amount of school-related reading										
Standardized coefficient	0.03	0.01	0.04 *	0.03	0.04 *	0.04 *	0.01	0.04 *	0.06 *	0.06*
Standard error	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
P-value	0.070	0.379	0.025	0.126	0.011	0.009	0.639	0.012	0.004	0.003
Amount of non-school-related reading										
Standardized coefficient	-0.01	0.01	0.00	-0.03	-0.02	-0.02	0.03	-0.03	-0.02	-0.03*
Standard error	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
P-value	0.518	0.434	0.845	0.071	0.253	0.220	0.108	0.117	0.229	0.040
Use of reflective reading strategies										
Standardized coefficient	0.04*	0.06 *	0.04 *	0.05 *	0.02	0.04 *	0.05 *	0.08 *	0.04 *	0.02
Standard error	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
P-value	0.002	0.000	0.007	0.000	0.092	0.017	0.003	0.000	0.028	0.197
F-test of reading achievement predictors	55.87*	36.87 *	35.98 *	36.77 *	30.70 *	31.86 *	19.47 *	46.48 *	14.34 *	12.33*
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F-test of reading behavior predictors	5.75*	8.90 *	7.47 *	6.42 *	4.06 *	5.65 *	6.34 *	12.10 *	5.95 *	4.37*
P-value	0.001	0.000	0.000	0.000	0.007	0.001	0.000	0.000	0.000	0.004
Sample size	4,293	4,252	3,955	4,129	4,221	3,891	3,848	3,588	3,518	3,783
R-squared	0.443	0.281	0.312	0.292	0.280	0.365	0.240	0.280	0.234	0.235

(continued)

Appendix Table J.1 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities school records data, follow-up GRADE assessment, and follow-up student survey.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The values in this table are standardized regression coefficients that measure the extent to which students' achievement and reading behaviors (at the end of the program year) predict their academic performance (at the end of the program year and follow-up year). Entries in the table represent the estimated effect of a 1 standard deviation increase in the reading predictor on academic performance, scaled as a proportion of the standard deviation in the academic performance measure (that is, as an effect size). Regression coefficients were obtained by fitting a model in which the academic performance measure is regressed against the reading predictors. All models control for the the blocking of random assignment by school and cohort, as well as the following covariates: whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, a baseline measurement of the outcome variable in the school year prior to ERO participation, and indicators of missing data for all relevant student characteristics.

Statistical significance of the coefficients is indicated (*) when the p-value is less than or equal to 5 percent.

The analyses are based on students in the school records sample who also have data on the reading predictors.

[a]Reading achievement is measured using the GRADE assessment, while reading behaviors are measured using the student survey. Both instruments were administered in spring 2005 for Cohort 1 and in spring 2006 for Cohort 2 (end of ninth grade).

The Enhanced Reading Opportunities Study

Appendix Table J.2

Association Between Reading Outcomes and Credit Accumulation in Core Subject Areas (Standardized Regression Coefficients)

Reading Predictors (End of Program Year)[a]	Core Credits Earned in the Program Year: All Core	ELA	Social Studies	Science	Math	Core Credits Earned in the Follow-Up Year: All Core	ELA	Social Studies	Science	Math
Reading achievement										
Reading comprehension										
Standardized coefficient	0.12*	0.07 *	0.08 *	0.09 *	0.11 *	0.10 *	0.07 *	0.09 *	0.07 *	0.09*
Standard error	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Reading vocabulary										
Standardized coefficient	0.01	0.01	0.01	0.01	-0.02	0.01	0.01	0.01	0.00	0.01
Standard error	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
P-value	0.654	0.743	0.661	0.594	0.363	0.447	0.791	0.600	0.983	0.706
Reading behaviors										
Amount of school-related reading										
Standardized coefficient	0.08*	0.03	0.03	0.06 *	0.08 *	0.09 *	0.04 *	0.06 *	0.06 *	0.09*
Standard error	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
P-value	0.000	0.093	0.067	0.001	0.000	0.000	0.022	0.000	0.000	0.000
Amount of non-school-related reading										
Standardized coefficient	-0.05*	-0.02	0.00	-0.06 *	-0.05 *	-0.05 *	-0.01	-0.04 *	-0.05 *	-0.05*
Standard error	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
P-value	0.003	0.286	0.889	0.001	0.007	0.001	0.756	0.014	0.008	0.008
Use of reflective reading strategies										
Standardized coefficient	0.04*	0.05 *	0.02	0.05 *	0.02	0.04 *	0.04 *	0.04 *	0.05 *	0.02
Standard error	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
P-value	0.008	0.004	0.132	0.001	0.284	0.010	0.038	0.008	0.001	0.331
F-test of reading achievement predictors	29.06*	9.27 *	14.49 *	15.80 *	19.12 *	21.49 *	9.19 *	17.77 *	9.72 *	17.16*
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F-test of reading behavior predictors	11.43*	5.02 *	3.33 *	9.65 *	7.35 *	15.20 *	5.17 *	8.60 *	10.09 *	9.20*
P-value	0.000	0.002	0.019	0.000	0.000	0.000	0.001	0.000	0.000	0.000
Sample size	4,293	4,293	4,293	4,293	4,293	3,891	3,891	3,891	3,891	3,891
R-squared	0.278	0.168	0.249	0.164	0.164	0.314	0.191	0.290	0.218	0.207

(continued)

Appendix Table J.2 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities school records data, follow-up GRADE assessment, and follow-up student survey.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The values in this table are standardized regression coefficients that measure the extent to which students' achievement and reading behaviors (at the end of the program year) predict their academic performance (at the end of the program year and follow-up year). Entries in the table represent the estimated effect of a 1 standard deviation increase in the reading predictor on academic performance, scaled as a proportion of the standard deviation in the academic performance measure (that is, as an effect size). Regression coefficients were obtained by fitting a model in which the academic performance measure is regressed against the reading predictors. All models control for the the blocking of random assignment by school and cohort, as well as the following covariates: whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, a baseline measurement of the outcome variable in the school year prior to ERO participation, and indicators of missing data for all relevant student characteristics.

Statistical significance of the coefficients is indicated (*) when the p-value is less than or equal to 5 percent.

The analyses are based on students in the school records sample who also have data on the reading predictors.

[a]Reading achievement is measured using the GRADE assessment, while reading behaviors are measured using the student survey. Both instruments were administered in spring 2005 for Cohort 1 and in spring 2006 for Cohort 2 (end of ninth grade).

The Enhanced Reading Opportunities Study

Appendix Table J.3

Association Between Reading Outcomes and Performance on State Tests (Standardized Regression Coefficients)

	State Test Scores in the Program Year				State Test Scores in the Follow-Up Year			
Reading Predictors (End of Program Year)[a]	ELA	Social Studies	Science	Math	ELA	Social Studies	Science	Math
Reading achievement								
Reading comprehension								
Standardized coefficient	0.20*	0.13 *	0.22 *	0.13 *	0.26 *	0.17*	0.17*	0.13*
Standard error	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
P-value	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Reading vocabulary								
Standardized coefficient	0.12*	0.20 *	0.06 *	0.04	0.04	0.05	0.09*	0.05
Standard error	(0.02)	(0.04)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
P-value	0.000	0.000	0.015	0.102	0.124	0.055	0.001	0.137
Reading behaviors								
Amount of school-related reading								
Standardized coefficient	0.01	0.02	0.07	0.02	0.05	0.00	0.03	0.07*
Standard error	(0.02)	(0.04)	(0.04)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
P-value	0.683	0.724	0.067	0.319	0.052	0.876	0.367	0.009
Amount of non-school-related reading								
Standardized coefficient	0.00	-0.02	0.01	-0.04	-0.03	0.02	0.01	-0.04
Standard error	(0.02)	(0.05)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
P-value	0.895	0.688	0.778	0.072	0.313	0.426	0.796	0.098
Use of reflective reading strategies								
Standardized coefficient	0.03	0.06	0.01	0.02	0.01	0.00	-0.04	-0.02
Standard error	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
P-value	0.174	0.081	0.585	0.392	0.667	0.993	0.072	0.359
F-test of reading achievement predictors	66.38*	25.49 *	35.11 *	23.24 *	68.22 *	30.54*	45.59*	19.31*
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F-test of reading behavior predictors	0.98	1.30	2.84 *	1.25	1.68	0.27	1.40	2.56
P-value	0.399	0.272	0.036	0.290	0.169	0.847	0.240	0.053
Sample size	2,024	839	2,118	2,380	2,199	2,066	2,456	2,317
R-squared	0.333	0.260	0.244	0.313	0.260	0.236	0.181	0.230

(continued)

Appendix Table J.3 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities school records data, follow-up GRADE assessment, and follow-up student survey.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The values in this table are standardized regression coefficients that measure the extent to which students' achievement and reading behaviors (at the end of the program year) predict their academic performance (at the end of the program year and follow-up year). Entries in the table represent the estimated effect of a 1 standard deviation increase in the reading predictor on academic performance, scaled as a proportion of the standard deviation in the academic performance measure (that is, as an effect size). Regression coefficients were obtained by fitting a model in which the academic performance measure is regressed against the reading predictors. All models control for the the blocking of random assignment by school and cohort, as well as the following covariates: whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, a baseline measurement of the outcome variable in the school year prior to ERO participation, and indicators of missing data for all relevant student characteristics.

Statistical significance of the coefficients is indicated (*) when the p-value is less than or equal to 5 percent.

The analyses are based on students in the state tests samples who also have data on the reading predictors.

[a]Reading achievement is measured using the GRADE assessment, while reading behaviors are measured using the student survey. Both instruments were administered in spring 2005 for Cohort 1 and in spring 2006 for Cohort 2 (end of ninth grade).

statistically significant. For instance in Table J.1, the magnitude of the association between GRADE scores and core GPA is 23 percent smaller in the follow-up year than in the program year, on average.

The associations presented in Appendix Tables J.1 to J.3 are regression-adjusted associations that control for the fact that the predictors (reading outcomes) are correlated among each other. In other words, the associations in these tables represent the association between a given reading outcome and students' academic performance, *controlling for the other reading outcomes*. In order to examine whether these adjustments affect the pattern of findings described above, simple bivariate Pearson correlations between the reading measures and the academic performance measures were also estimated.[5] These unadjusted correlations are presented in Appendix Tables J.4 to J.6; the values of the correlations range from 0 to 1 (no correlation to perfect correlation).[6] In general, the pattern of findings for the bivariate correlations is similar to the pattern described above for the regression-adjusted associations, with one exception:

- Based on simple bivariate correlations (Tables J.4 to J.6), students' GRADE vocabulary scores are correlated with their performance in core courses and on state tests by a statistically significant amount (correlations are at least 0.10). However, when these correlations are regression-adjusted for other reading outcomes (that is, reading comprehension scores and reading behaviors), the association between students' vocabulary scores and their academic outcomes is no longer statistically significant (Tables J.1 to J.3). This is due to the fact that GRADE reading comprehension scores and vocabulary scores are correlated. Therefore, once reading comprehension scores have been taken into account, vocabulary scores are not a statistically significant predictor of academic performance.

[5]The analyses are based on students in the school records sample who also have data on the relevant reading outcome.

[6]Because reading outcomes and the academic performance outcomes are both measured with error, the values in these tables are estimates of the correlation between these two types of outcome, and hence the table also presents tests of the hypothesis that these correlations are equal to zero.

The Enhanced Reading Opportunities Study

Appendix Table J.4

Correlations Between Reading Outcomes and Grade Point Average in Core Subject Areas

Reading Outcomes (End of Program Year)[a]	Core GPA in the Program Year: All Core	ELA	Social Studies	Science	Math	Core GPA in the Follow-Up Year: All Core	ELA	Social Studies	Science	Math
Reading achievement										
Reading comprehension										
Correlation coefficient	0.27*	0.24*	0.22*	0.22*	0.20*	0.23*	0.19*	0.22*	0.18*	0.15*
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sample size	4,456	4,411	4,088	4,280	4,375	4,028	3,983	3,717	3,645	3,917
Reading vocabulary										
Correlation coefficient	0.13*	0.11*	0.12*	0.10*	0.09*	0.13*	0.11*	0.13*	0.10*	0.10*
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sample size	4,456	4,411	4,088	4,280	4,375	4,028	3,983	3,717	3,645	3,917
Reading behaviors										
Amount of school-related reading										
Correlation coefficient	0.10*	0.07*	0.10*	0.08*	0.08*	0.11*	0.09*	0.11*	0.10*	0.08*
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sample size	4,344	4,301	3,996	4,179	4,270	3,935	3,891	3,630	3,557	3,826
Amount of non-school-related reading										
Correlation coefficient	0.03*	0.04*	0.05*	0.01	0.01	0.04*	0.06*	0.03*	0.03	0.01
P-value	0.034	0.014	0.004	0.427	0.540	0.013	0.000	0.039	0.059	0.730
Sample size	4,395	4,351	4,036	4,223	4,315	3,980	3,935	3,674	3,602	3,870
Use of reflective reading strategies										
Correlation coefficient	0.10*	0.09*	0.08*	0.09*	0.06*	0.10*	0.09*	0.12*	0.09*	0.05*
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003
Sample size	4,325	4,283	3,982	4,161	4,252	3,921	3,878	3,616	3,544	3,813

(continued)

Appendix Table J.4 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities school records data, follow-up GRADE assessment, and follow-up student survey.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The values in this table are bivariate Pearson correlation coefficients that measure the association between each reading measure (at the end of the program year) and each academic performance measure (at the end of the program year and follow-up year). Values range from 0 to 1. Statistical significance of the coefficients is indicated (*) when the p-value is less than or equal to 5 percent.

The analyses are based on students in the school records sample who also have data on the relevant reading measure.

[a]Reading achievement is measured using the GRADE assessment, while reading behaviors are measured using the student survey. Both instruments were administered in spring 2005 for Cohort 1 and in spring 2006 for Cohort 2 (end of ninth grade).

The Enhanced Reading Opportunities Study

Appendix Table J.5

Correlations Between Reading Outcomes and Credit Accumulation in Core Subject Areas

Reading Outcomes (End of Program Year)[a]	Core Credits Earned in the Program Year					Core Credits Earned in the Follow-Up Year				
	All Core	ELA	Social Studies	Science	Math	All Core	ELA	Social Studies	Science	Math
Reading achievement										
Reading comprehension										
Correlation coefficient	0.23*	0.15*	0.16*	0.16*	0.17*	0.21*	0.15*	0.16*	0.14*	0.19*
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sample size	4,456	4,456	4,456	4,456	4,456	4,028	4,028	4,028	4,028	4,028
Reading vocabulary										
Correlation coefficient	0.14*	0.09*	0.13*	0.09*	0.07*	0.15*	0.12*	0.14*	0.10*	0.10*
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sample size	4,456	4,456	4,456	4,456	4,456	4,028	4,028	4,028	4,028	4,028
Reading behaviors										
Amount of school-related reading										
Correlation coefficient	0.07*	0.03	0.08*	0.03*	0.06*	0.09*	0.05*	0.10*	0.03*	0.08*
P-value	0.000	0.078	0.000	0.037	0.000	0.000	0.002	0.000	0.046	0.000
Sample size	4,344	4,344	4,344	4,344	4,344	3,935	3,935	3,935	3,935	3,935
Amount of non-school-related reading										
Correlation coefficient	0.00	0.00	0.04*	-0.03	-0.02	0.00	0.02	0.03	-0.03	0.00
P-value	0.936	0.838	0.003	0.056	0.311	0.821	0.178	0.089	0.070	0.841
Sample size	4,395	4,395	4,395	4,395	4,395	3,980	3,980	3,980	3,980	3,980
Use of reflective reading strategies										
Correlation coefficient	0.07*	0.05*	0.08*	0.05*	0.04*	0.07*	0.06*	0.09*	0.04*	0.05*
P-value	0.000	0.001	0.000	0.002	0.010	0.000	0.000	0.000	0.027	0.005
Sample size	4,325	4,325	4,325	4,325	4,325	3,921	3,921	3,921	3,921	3,921

(continued)

Appendix Table J.5 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities school records data, follow-up GRADE assessment, and follow-up student survey.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The values in this table are bivariate Pearson correlation coefficients that measure the association between each reading measure (at the end of the program year) and each academic performance measure (at the end of the program year and follow-up year). Values range from 0 to 1. Statistical significance of the coefficients is indicated (*) when the p-value is less than or equal to 5 percent.

The analyses are based on students in the school records sample who also have data on the relevant reading measure.

[a]Reading achievement is measured using the GRADE assessment, while reading behaviors are measured using the student survey. Both instruments were administered in spring 2005 for Cohort 1 and in spring 2006 for Cohort 2 (end of ninth grade).

The Enhanced Reading Opportunities Study

Appendix Table J.6

Correlations Between Reading Outcomes and Performance on State Tests

Reading Outcomes (End of Program Year)[a]	State Test Scores in the Program Year				State Test Scores in the Follow-Up Year			
	ELA	Social Studies	Science	Math	ELA	Social Studies	Science	Math
Reading achievement								
Reading comprehension								
Correlation coefficient	0.43*	0.31*	0.33*	0.27*	0.38*	0.32*	0.28*	0.23*
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sample size	2,077	879	2,186	2,473	2,271	2,143	2,529	2,380
Reading vocabulary								
Correlation coefficient	0.35*	0.32*	0.25*	0.21*	0.25*	0.24*	0.22*	0.18*
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sample size	2,077	879	2,186	2,473	2,271	2,143	2,529	2,380
Reading behaviors								
Amount of school-related reading								
Correlation coefficient	0.07*	0.01	0.08*	0.03	0.07*	0.02	0.02	0.03
P-value	0.001	0.819	0.000	0.148	0.001	0.477	0.447	0.173
Sample size	2,047	851	2,140	2,402	2,223	2,087	2,477	2,339
Amount of non-school-related reading								
Correlation coefficient	0.02	-0.01	0.04	-0.02	0.00	0.00	-0.01	-0.03
P-value	0.371	0.750	0.083	0.293	0.853	0.947	0.514	0.167
Sample size	2,057	866	2,162	2,441	2,251	2,119	2,503	2,357
Use of reflective reading strategies								
Correlation coefficient	0.08*	0.07*	0.05*	0.03	0.04*	0.01	-0.02	0.00
P-value	0.001	0.035	0.023	0.196	0.036	0.711	0.266	0.829
Sample size	2,033	847	2,131	2,394	2,211	2,084	2,474	2,332

(continued)

Appendix Table J.6 (continued)

SOURCES: MDRC calculations from the Enhanced Reading Opportunities school records data, follow-up GRADE assessment, and follow-up student survey.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The values in this table are bivariate Pearson correlation coefficients that measure the association between each reading measure (at the end of the program year) and each academic performance measure (at the end of the program year and follow-up year). Values range from 0 to 1. Statistical significance of the coefficients is indicated (*) when the p-value is less than or equal to 5 percent.

The analyses are based on students in the state test samples who also have data on the relevant reading measure.

[a]Reading achievement is measured using the GRADE assessment, while reading behaviors are measured using the student survey. Both instruments were administered in spring 2005 for Cohort 1 and in spring 2006 for Cohort 2 (end of ninth grade).

Appendix K

Variation in Impacts Across Sites and Cohorts

This appendix provides supplemental information on the association between implementation characteristics and the impacts of the Enhanced Reading Opportunities (ERO) programs (Chapter 5). The first section of this appendix presents impact estimates on course performance outcomes (grade point average [GPA] and credit accumulation) for each school in the study. It also describes the statistical model used to estimate these impacts, as well as tests of site variation that were conducted. The second section discusses the statistical model used to estimate the association between impacts and implementation characteristics (presented in Chapter 5), and includes additional findings on these associations that are not included in that chapter.

Impact Estimates by School

This section presents observed impact estimates for each random assignment block. These estimates are obtained by fitting the following model to the school records samples:

$$Y_{it} = \sum_{K} \lambda_{kt} B_{ki} + \sum_{J} \phi_{jt} T_i * S_{ji} + \sum_{S} \delta_s X_{si} + \varepsilon_{it} \qquad (1)$$

where:

Y_{it} = Outcome of interest for student i in year t (program year or follow-up year).

T_i = Indicator of ERO group membership (treatment status). This indicator is equal to 1 if student i was assigned to the ERO program and zero otherwise.

B_{ki} = Random assignment block indicators (school-by-cohort), equal to 1 if student i is in random assignment block k and zero otherwise.[1]

S_{ki} = School indicators, equal to 1 if student i is in school j and zero otherwise.[2]

X_{si} = Set of S pre-random assignment characteristics and prior achievement outcomes for student i. These are the same characteristics that are included in the impact model to improve the precision of the impact estimates (see Appendix F).

ε_{it} = A within-student error term.

[1]There are 68 random assignment block indicators (34 schools * 2 implementation years).
[2]There are 34 school indicators.

Therefore:

ϕ_{jt} = The observed (fixed-effects) impact estimate for school *j* in follow-up year *t*.

Robust standard errors are used when testing whether the estimated impact of the program is equal to zero.[3]

Figures 5.1 and 5.2 present estimated program impacts on GPA in core subject areas for each school, while Appendix Figures K.1 and K.2 present school-specific estimated impacts on credit accumulation.

Test of Variation in Impacts

Because each impact is estimated with error, the range of impacts in these figures overstates the true amount of variation in impacts across schools. Therefore, statistical tests were conducted to determine whether the variation in observed impacts across schools is statistically significant, given noise and sampling error. These tests are conducted by fitting the following model:

$$Y_{it} = \alpha_t + \sum_K \lambda_{kt} B_{ki} + \beta_t T_t + \sum_J \phi_{jt} T_i * S_{ji} + \sum_S \delta_s X_{si} + \varepsilon_{it} \qquad (2)$$

and testing whether the estimates of ϕ_{jt} are jointly equal to zero, based on an omnibus F-test. The results are as follows:

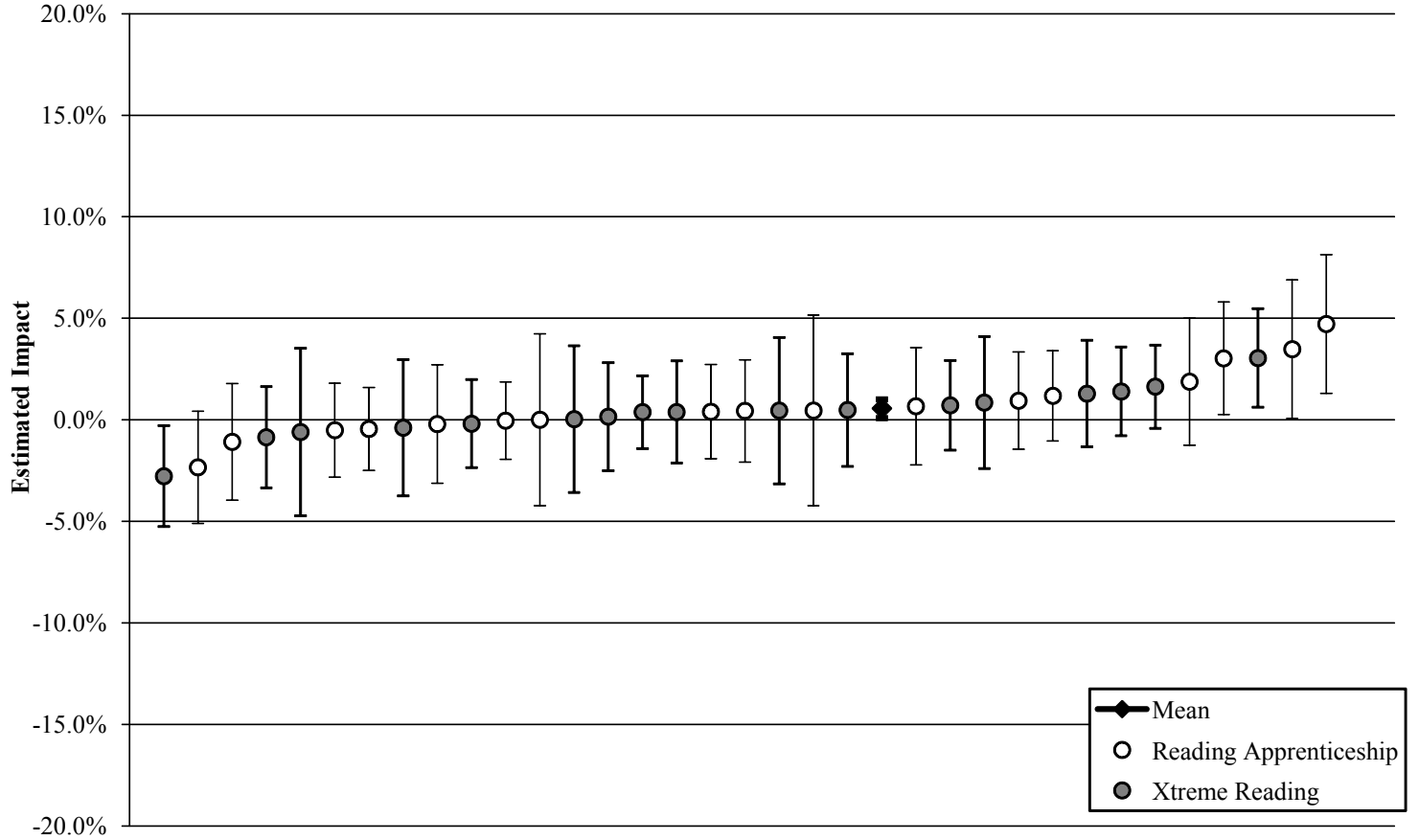
- In the program year (ninth grade), variation in observed impact estimates across schools is not statistically significant at the 5 percent level for GPA or credit accumulation in core subject areas (p-value = 0.101 for GPA and 0.346 for credit accumulation).
- In the follow-up year (tenth grade), the variation in observed impact estimates on core GPA is statistically significant (p-value = 0.008). The variation in impacts on core credit accumulation is not statistically significant (p-value = 0.535).

[3]Huber-White standard errors are used to account for heteroskedasticity, which in this case may result from the ERO programs having an impact on the variance of the outcome (that is, the variance in *Y* may differ for the ERO and non-ERO group as a result of the program).

The Enhanced Reading Opportunities Study

Appendix Figure K.1

Fixed-Effect Impact Estimates on Credit Accumulation in Core Subject Areas in the Program Year, by School, School Records Sample



(continued)

Appendix Figure K.1 (continued)

SOURCE: MDRC calculations from individual students' school records data supplied by each school district.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2.

The fixed-effects impact estimates in this figure are the regression-adjusted coefficients on the interaction between schools and the treatment indicator. These estimates are adjusted for random baseline differences between the ERO and the non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation.

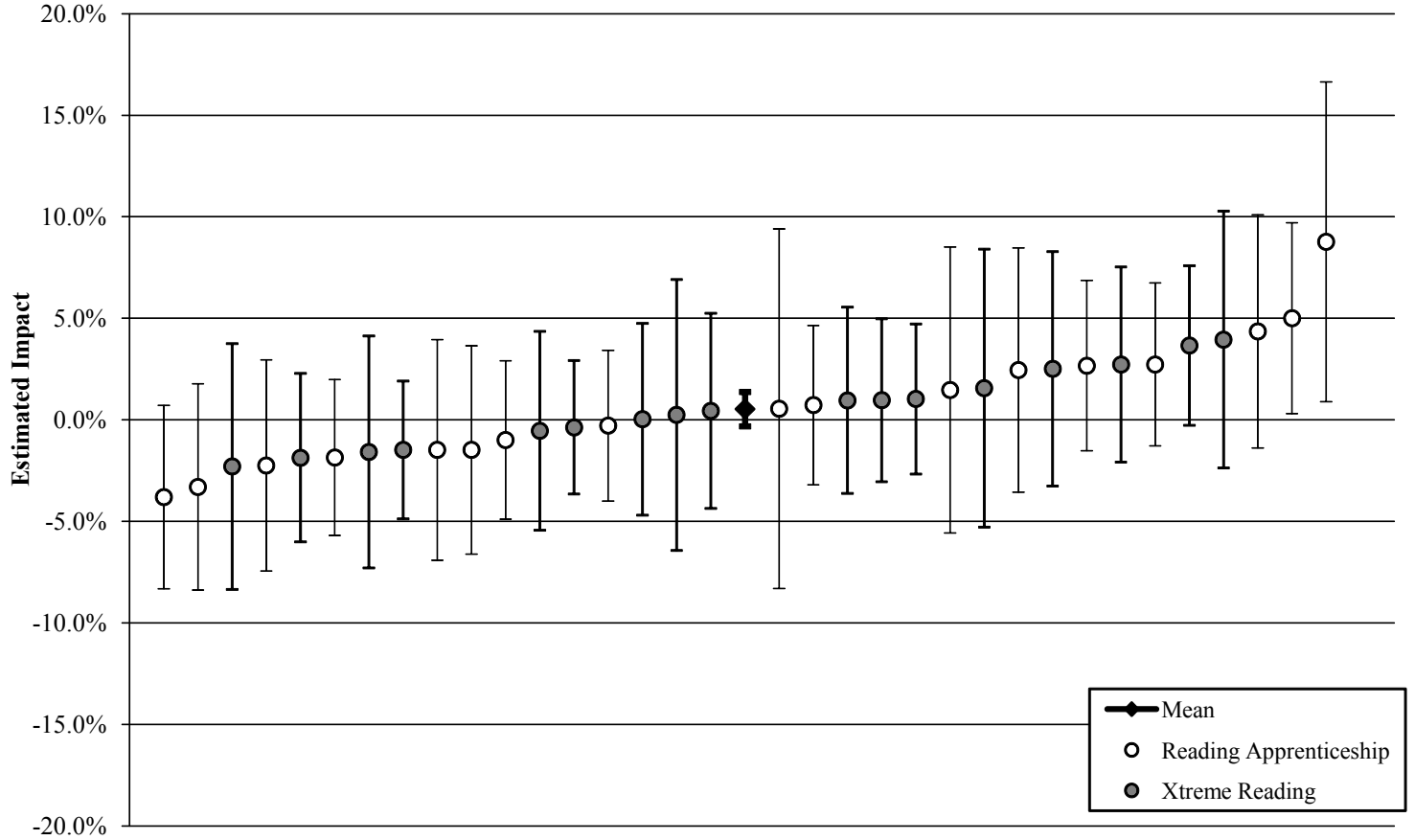
Credit accumulation is defined as the cumulative number of credits earned, scaled as a percentage of the total number of core credits required for graduation in a student's district.

Sample size: 5,073 students.

The Enhanced Reading Opportunities Study

Appendix Figure K.2

Fixed-Effect Impact Estimates on Credit Accumulation in Core Subject Areas in the Follow-Up Year, by School, School Records Sample



(continued)

Appendix Figure K.2 (continued)

SOURCE: MDRC calculations from individual students' school records data supplied by each school district.

NOTES: The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The fixed-effects impact estimates in this figure are the regression-adjusted coefficients on the interaction between schools and the treatment indicator. These estimates are adjusted for random baseline differences between the ERO and the non-ERO groups in terms of the following variables: GRADE reading comprehension test score at baseline, whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, and a baseline measurement of the outcome variable in the school year prior to ERO participation.

Credit accumulation is defined as the cumulative number of credits earned, scaled as a percentage of the total number of core credits required for graduation in a student's district.

Sample size: 4,359 students.

Overall, these findings indicate that in the ninth grade, the programs' effect on academic performance does not differ across the study schools. However, in the follow-up year (tenth grade), the program's effect on GPA is larger in some schools than others.

Amount of Variation in Impacts (Noise-Corrected)

In order to estimate the amount of true variation in impacts after correcting for noise (which are reported in Chapter 5), the following model was estimated based on the school records samples:

$$Y_{it} = \beta_t T_i + \sum_K \lambda_{kt} B_{ki} + \sum_S \delta_s X_{si} + \mu_{jt} * T_i + \varepsilon_{it} \quad (3)$$

where variables are defined as before, and:

β_t = The average (random effects) estimate impact of the ERO programs on outcome Y in follow-up year t, pooled across all random assignment blocks.

The centerpiece of this model is:

μ_{jt} = A between-school error term for students in the ERO group. This error term represents the *difference* between the impact of the program in school j and the average (pooled) impact of the programs across all blocks (β_t), adjusted for estimation error, in year t.[4]

The variation in μ_{jt} represents the noise-corrected variation in impacts across blocks.

Associations Between Impacts and Implementation Characteristics

This section describes the statistical model used to estimate the association between program impacts and various features of program implementation — including the fidelity with which the programs were implemented, as well as the literacy environment of the study schools. This section also presents findings on the association between these implementation characteris-

[4]For readers familiar with multilevel modeling, Equation (3) is a multilevel model with a random treatment effect (slope). Because it has two error terms, this model must be estimated using maximum likelihood. (Ordinary least squares regression minimizes a simple "one component" residual and cannot minimize a two-part residual.)

tics and program impacts on the second primary school outcome in this study: credit accumulation. When interpreting these findings, it is important to remember that associations between impacts and implementation characteristics do not necessarily represent true causal relationships, because these characteristics were not randomly assigned to the study schools.

Statistical Model

The association between implementation characteristics and program impacts on course performance was estimated by fitting the following model to the school records samples:

$$Y_{it} = \sum_{K} \lambda_{kt} B_{ki} + \theta_{1t} T_i * FID_{ki} + \theta_{2t} T_i * LITCL_{ki} + \theta_{3t} T_i * LITTUT_{ki} + \sum_{S} \delta_s X_{si} + \varepsilon_{it} \quad (4)$$

where variables are defined as before, and where the implementation characteristics of interest are:

FID_{ki} = Average composite implementation fidelity rating for random assignment block k (school-by-cohort indicator).[5]

$LITCL_{ki}$ = Average number of school-based literacy class sessions taken by the non-ERO group in random assignment block k during the program year.[6]

$LITTUT_{ki}$ = Average number of school-based tutoring sessions taken by the non-ERO group in random assignment block k during the program year.

Therefore:

θ_{1t} = The estimated association between impacts on outcome Y and average implementation fidelity ratings.

θ_{2t} = The estimated association between impacts on outcome Y and the average the number of literacy classes taken by students in the non-ERO group.

θ_{3t} = The estimated association between impacts on outcome Y and the average the number of tutoring sessions taken by students in the non-ERO group.

---

[5]The ratings are from classroom observations and are based on a scale of 1 to 3. Appendix C provides further details on the measurement of implementation fidelity in the ERO study.

[6]Appendix B provides further detail on these measures of supplemental literacy services, which are from the student follow-up survey.

Robust standard errors were used when testing whether these estimated associations are equal to zero.

Association Between Implementation Characteristics and Impacts on Credit Accumulation

Appendix Table K.1 presents estimated associations between impacts on credit accumulation and the three implementation characteristics listed above.[7] The values in Table K.1 are standardized regression coefficients. That is, they represent the change in the impact on credits earned that is predicted by a 1 standard deviation increase in a particular implementation characteristic, scaled as an effect size (that is, as a proportion of the standard deviation in credit accumulation). As seen in this table:

- Estimated impacts on credit accumulation are not associated with either the fidelity with which the program was implemented or the school's literacy environment.

[7]These findings were not presented in Chapter 5, which focuses on the association between implementation characteristics and impacts on students' GPA rather than their credit accumulation.

The Enhanced Reading Opportunities Study

Appendix Table K.1

Association Between Program Implementation Conditions and Impacts on Credit Accumulation in Core Subject Areas (Standardized Regression Coefficients)

Implementation Conditions (Program Year)	Association with Impacts on Credit Accumulation in the: Program Year	Follow-Up Year
ERO program fidelity[a]		
Composite implementation fidelity rating		
Estimated coefficient	0.01	0.00
Standard error	(0.02)	(0.03)
P-value	0.704	0.932
Non-ERO supplemental literacy support services[b]		
Number of school-based literacy classes		
Estimated coefficient	-0.03	0.01
Standard error	(0.03)	(0.03)
P-value	0.237	0.679
Number of tutoring sessions with school-based adult tutor		
Estimated coefficient	0.02	0.00
Standard error	(0.03)	(0.03)
P-value	0.602	0.955
Sample size	5,150	4,436
R-squared	0.282	0.306

SOURCES: MDRC calculations from the Enhanced Reading Opportunities school records data, classroom observations, and follow-up survey.

NOTES: The program year is the year in which students were enrolled in an ERO class; it corresponds to the 2005-2006 school year for Cohort 1 and the 2006-2007 school year for Cohort 2. The follow-up year corresponds to the 2006-2007 school year for Cohort 1 and the 2007-2008 school year for Cohort 2.

The values in this table are standardized regression coefficients that measure the extent to which the implementation conditions in the study schools are associated with impacts on academic performance in the program year and follow-up year. Entries in the table represent the estimated association between a 1 standard deviation increase in the implementation measure and the magnitude of the impact of the ERO programs on academic performance; these associations are scaled as a proportion of the standard deviation in the academic performance measure for the non-ERO group (that is, as effect sizes).

Regression coefficients were obtained by fitting an impact model that includes an indicator of treatment status, as well as a set of interactions between the treatment indicator and the implementation conditions; the values reported in this table are the coefficients on these interaction terms. All models also control for the blocking of random assignment by school and cohort, as well as the following covariates: whether a student was overage for grade at random assignment, a student's score on the standardized reading and math assessments (in standardized units) administered by the school district in the year prior to ERO participation, a baseline measurement of the outcome variable in the school year prior to ERO participation, and indicators of missing data for all relevant student characteristics.

Statistical significance of the coefficients is indicated (*) when the p-value is less than or equal to 5 percent.

[a]Fidelity ratings are based on classroom observations conducted in the spring of the program year (spring 2005 for Cohort 1 and spring 2006 for Cohort 2). The composite rating captures implementation fidelity on two key dimensions: the classroom learning environment and the teacher's use of instructional strategies focused on reading comprehension. The composite rating is on a scale of 1 to 3.

[b]These measures are from the follow-up student survey, administered at the end of the program year (spring 2005 for Cohort 1 and spring 2006 for Cohort 2). They are based on the average number of literacy classes or tutoring sessions taken by students in the non-ERO group in each school and cohort during the program year.

Appendix L

Program Costs

This appendix provides additional information that is relevant to the Enhanced Reading Opportunities (ERO) cost study discussed in Chapter 6. The first section of the appendix presents a more detailed breakdown of the nonsalary program costs that were used to calculate the costs presented in Chapter 6. The second section discusses the extent to which there is variation in program costs across the study schools and examines the two factors that are the most important contributors of this variation (teacher characteristics and the number of students served). The third section concludes by providing further detail on the calculation of program costs in nonstudy settings.

The Resource Cost Model (RCM) Approach

The cost analysis in Chapter 6 uses the RCM approach[1] to calculate the costs of delivering the ERO programs at the study sites. The approach combines information on the quantities of personnel and resources with their unit costs (for example, teacher yearly salary and fringes, program materials) to produce an estimate of overall costs. By breaking down costs and quantities of resources in this way, the RCM approach allows both for identification of factors that have the greatest impact on costs and also for simulation of how total costs change in situations where different levels of resources are used or unit costs vary. This flexibility allows the research team to address questions of overall costs in nonstudy settings and to understand how these costs would change under alternative assumptions.

Data Sources

Several data sources were used to calculate program costs. These include documents prepared by districts as part of the application for their Smaller Learning Community (SLC) grants; predetermined cost schedules presented to participating districts and schools; developer budget documents presented to American Institutes for Research (AIR); internal AIR cost records; and publicly available resources such as teacher salary step schedules downloaded from district Web sites[2] and the National Center for Education Statistics (NCES) Common Core

[1]Levin, Catlin, and Elson (2007); Levin, Catlin, and Elson (2009); Levin and McEwan (2001).

[2]One district did not have a teacher step salary schedule available, but instead posted information on the average teacher salary at each of its ERO schools, which were used as proxies. In some districts the teacher step salary schedule was not always available for the two study years (2005-2006 and 2006-2007). In these cases, salaries were adjusted accordingly for each study year, using information from the Bureau of Labor Statistics Employment Cost Index (ECI) for Elementary and Secondary Schools.

of Data (CCD) Fiscal Survey (F-33 Data).[3] Cost data were collected for the two years of program implementation (2005-2006 and 2006-2007).

Adjustments to Cost Data

The cost data available for this study are for the two years of program implementation, 2005-2006 and 2006-2007. Two adjustments were made to make the costs more informative. Initial cost figures for each school (by cost category and program year) were transformed for each program year into *real* 2008 dollars. Personal costs were converted using adjustments derived from the Bureau of Labor Statistics (BLS) Employment Cost Index for Elementary and Secondary Schools. Nonpersonnel costs were adjusted for inflation in a similar manner, using a schedule derived from the Consumer Price Index (CPI) for all Urban Consumers.[4]

Personnel costs were also adjusted to reflect variations in the price level of personnel resources across the geographic areas in which the study took place. These adjustments were made using the National Center for Education Statistics (NCES) Comparable Wage Index (CWI).[5] Nonpersonnel costs were not adjusted to reflect geographic differences because they had been predetermined as a study constraint.

Analysis Procedures

Once total costs for each of the components in the RCM were derived, amounts were converted to per-student costs for each site. The total number of students enrolled in demonstra-

---

[3]Yearly benefit rates were calculated on estimated salaries for each study district using the NCES Common Core of Data (CCD) Fiscal Survey (F-33 Data). The F-33 data provide information on the total dollars spent on both direct salaries and benefits for instructional staff. District benefit rates were calculated by dividing the total dollars spent on instructional staff benefits by the total dollars spent on instructional salaries within each district. The F-33 data and corresponding documentation are accessible at http://nces.ed.gov/ccd/f33agency.asp.

For project coordinator salaries, only one district SLC proposal provided salary information for the project coordinator; based on this information, project coordinator salaries for other districts were estimated using the National Center for Education Statistics (NCES) Comparable Wage Index (CWI).

[4]Personnel and fringe benefit dollars for Year 1 (2005-2006) and Year 2 (2006-2007) were inflated using adjustment factors of 1.114 and 1.069, respectively; nonpersonnel dollars for Year 1 (2005-2006) and Year 2 (2006-2007) were inflated using adjustment factor of 1.102 and 1.068, respectively.

[5]The CWI index values reflect the cost of hiring and retaining college graduate staff in each district *relative* to the average district in the nation as a whole. The index is based on estimations of the average regional labor market wage earned by noneducators whose occupations require a similar level of training to that of teachers. For a detailed discussion of the CWI, see Taylor and Fowler (2005). The CWI data files and corresponding documentation are available for download at https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007397.

tion schools ranged from 32 to 72.[6] The range reflected both the number of eligible students who could be recruited and the success of the recruitment efforts. Average per-student costs across all schools were obtained by summing costs across all students and dividing by the total number of students.

Detailed Nonsalary Costs

Appendix Table L.1 presents the nonsalary costs *per school* of the ERO programs (training, travel, equipment/supplies, other direct costs, and indirect costs) that were used to calculate the costs reported in Chapter 6. The first panel of Table L.1 presents costs associated with the summer institute and follow-up training for ERO teachers, while the second panel of the table presents costs associated with implementing the programs during the school year (implementation support, professional development, and materials).[7]

Note that the training costs in Appendix Table L.1 include the cost of training replacement teachers in the 10 schools that experienced teacher attrition during the study.[8] These schools incurred the additional cost of providing their replacement teacher with training from the program developers, which varied depending on whether the school was implementing Xtreme Reading or Reading Apprenticeship Academic Literacy (RAAL). The Xtreme Reading replacement teacher training cost was $4,372 in each of the five schools, while the RAAL costs averaged $2,582 over the five schools. In both cases, these costs consisted of developer charges, lodging, food, participant per-diem costs, and coordination charges.[9] Note that the cost of training replacement teachers is not included in the cost simulations for nonstudy settings (Table 6.2).

---

[6]This range reflects students actually enrolled in the ERO classes. Under ideal study conditions, each ERO class would enroll 15 students, and each teacher would teach four sections. When recruitment efforts failed to meet their targets, and it was impossible to create four treatment classes with a comparable number of comparison students, fewer ERO sessions were offered and teachers had other assignments.

[7]For training and travel costs, total costs per school are the sum of the training or travel cost in the first panel ("Summer Institute and Follow-Up Training") and the second panel ("Implementation Support, Professional Development, and Materials"). For example, Year 1 training costs are $5,015 + 9,993, or $15,000.

[8]The discussions of returning and replacement teachers in the second report (Corrin et al., 2008) and earlier in this report (for example, Chapter 3) identify nine replacement teachers, not 10. The distinction is that one teacher withdrew from teaching the ERO program, and a replacement was assigned after the conclusion of the first summer training but before the start of the ERO class. This teacher received training separately, provided by the program developer but paid for by the district. Thus, this teacher replacement has implications for the cost calculations but does not represent a mid-implementation instructional transition for students.

[9]Training costs did not vary across the 24 schools where the ERO teacher stayed on for both years of the study (since replacement teacher training costs were not incurred).

The Enhanced Reading Opportunities Study

Appendix Table L.1

Nonsalary Costs, by Implementation Year and Cost Category (in Nominal Dollars)

Cost Category	Description	Year 1 ($)	Year 2 ($)	Total ($)
Summer Institute and Follow-Up Training - Coordination and Logistics[a]				
Training	Conference Coordination/Faculty Usage Meeting - Includes room rental, A/V equipment, hotel fees and coordination staff time	1,565	1,565	3,130
	Food - Includes breakfast, lunch and refreshments	1,506	1,506	3,012
	Lodging - Hotel rooms	1,944	1,944	3,888
Subtotal	Total Summer Institute/Follow-up Training	5,015	5,015	10,030
Travel	Airfare and transfers for one teacher per school and one administrator per district	2,485	2,485	4,970
Subtotal Per School		7,500	7,500	15,000
Implementation Support, Professional Development, and Materials				
Training	Developer staff, consultants and coaches	9,993	6,083	16,076
Travel	For summer institutes, on-site monitoring of program implementation, follow-up trainings	4,369	2,549	6,918
Equipment/Supplies	Training materials, curriculum guides, workbooks, textbooks, computer software and and licenses, anthologies, instructional kits, etc.	4,100	1,796	5,896
Other Direct Costs	Communications, printing, photocopying, technical support, etc.	2,060	1,846	3,906
Indirect Costs		4,478	2,726	7,204
Subtotal Per School		25,000	15,000	40,000
Total Cost Per School		32,500	22,500	55,000

SOURCE: Statement of Roles and Responsibilities for Participation in the Enhanced Reading Opportunities Study - Estimated Costs of Services and Materials provided by the program developers to AIR.

NOTE:

[a]Includes expenses for one 5-day Summer Training Institute and up to two 2-day follow-up training institutes each year. Costs assume that one teacher per school and one district administrator or coordinator will attend each training institute. Travel costs may vary by school district. Year 2 costs assume that location, travel, and facility usage will be the same as for Year 1.

Factors Explaining Variation in Per-Student Costs in the Study Setting

As discussed in Chapter 6, the per-student cost of implementing the ERO programs differs across the schools in the study. Exploring the factors that drive this variation can provide useful information to practitioners who are considering whether to implement the reading interventions. As discussed in Chapter 6, variation in program costs across the study schools appears to be driven by two factors: the credentials of the ERO teacher (which affects personnel costs), and the number of students served by the ERO program (since higher program enrollment reduces the amount spent per student, given fixed costs).

Appendix Table L.2 examines the relationship between costs, teacher credentials, and program enrollment more systematically. For this part of the analysis, the study schools were grouped into three categories based on their per-student program costs: the 11 highest spending schools, the 11 lowest spending schools, and the 12 middle spending schools.[10] For each of these three groups, the table presents the per-student cost of the ERO programs (first column), the average number of students enrolled in the ERO program (second column), the ERO teacher's experience (third column), and his or her educational attainment (fourth to sixth columns).

Table L.2 indicates that the qualifications of the instructors selected to teach the ERO classes are an important consideration that will drive the per-student cost of implementing this type of program:[11]

- On average, schools that spent the most per student had a more experienced ERO teacher. The ERO teacher in the highest-spending schools has 14 years of experience, while in the lowest-spending schools ERO teachers had half of this experience level on average.[12]

- On average, schools that spent the most per student also had ERO teachers with higher educational attainment. Approximately 82 percent of teachers in the highest spending schools had a master's degree or above, whereas 59

[10]Grouping the schools into these spending categories is a means of avoiding outliers, which in turn makes it easier to see the relationship between schools' resource decisions and subsequent program costs.

[11]The difference in the average per-student program cost between the highest- and lowest-spending schools is driven primarily by costs associated with personnel and fringe benefit resources (specifically, the difference in personnel and fringe costs per student between the highest- and lowest-spending schools accounts for 87.4 percent of the total per-student cost difference between the highest- and lowest-spending schools).

[12]ERO teachers' experience ranged from one to 34 years.

The Enhanced Reading Opportunities Study

Appendix Table L.2

Average Cost per Student, Number of Students Served, and Teacher Qualifications in Highest-, Middle-, and Lowest-Spending Program Demonstration Schools (Figures Calculated Over Both Implementation Years)

School Category	Average Cost per Student ($)	Average Number of Students Enrolled Across Study Years	Average Years of Teacher Experience	Percentage of Teachers with BA	Percentage of Teachers with MA	Percentage of Teachers with Education Specialist or PhD Degree
All schools	1,931	57	11.4	27.9	55.9	16.2
Highest-spending schools	2,357	52	14	18.2	68.2	13.6
Middle-spending schools	1,897	57	12.9	25.0	45.8	29.2
Lowest-spending schools	1,607	61	7	40.9	54.5	4.5

SOURCES: AIR calculations derived from the school-level budget blueprint provided in the Statement of Roles and Responsibilities for Participation in the Enhanced Reading Opportunities Study, salary step schedules downloaded from district Web sites, teacher qualifications from the Enhanced Reading Opportunities teacher survey, the National Center for Education Statistics (NCES) Common Core of Data (CCD) Fiscal Survey (F-33 Data), NCES CWI data, BLS Employment Cost Index for Elementary and Secondary Schools, CPI for all Urban Consumers, and Internal AIR Records of Costs of Training Replacement Teachers.

NOTE: Rounding may cause slight discrepancies in calculating sums and differences.

percent of those teachers in low-spending schools had similar education levels.

- On average, schools that spent the most per student enrolled fewer students in the ERO program. In particular, the average number of students served ranged from 52 in the highest-spending schools (or an average class size of 13 students per section) to 61 in the lowest-spending schools (or average class size of 15.3 students).

Cost Calculations for Nonstudy Setting

The last section of Chapter 6 presents cost simulations under different assumptions about the implementation of the ERO programs in a nonstudy setting. The following additional information is provided about these assumptions:

- **Teacher salaries:** Table L.3 presents the salary estimates used to calculate the simulated per-student cost of the programs in Chapter 6 (Table 6.2).

The Enhanced Reading Opportunities Study

Appendix Table L.3

Teacher Salary, by Years of Experience and Highest Degree Earned (in 2008-2009 Dollars)

	0 – 1 Year of Experience		10 Years of Experience		20 Years of Experience	
Degree	Salary ($)	Total Compensation ($)	Salary ($)	Total Compensation ($)	Salary ($)	Total Compensation ($)
Bachelor's	39,811	51,855	47,766	62,215	55,786	72,662
Master's	42,824	55,779	51,314	66,837	61,541	80,157
Doctorate	47,161	61,428	57,947	75,476	70,211	91,450

SOURCES: AIR calculations derived from salary step schedules downloaded from district Web sites and the National Center for Education Statistics (NCES) Common Core of Data (CCD) Fiscal Survey (F-33 Data) and NCES CWI data.

NOTES: Figures represent average annual salary and fringe benefits, by years of experience and highest degree obtained. Total compensation equals salary plus fringe benefits. Fringe benefits are calculated by multiplying salary by the average fringe benefit rate across the ERO demonstration districts, of 30.3 percent.

- **Nonsalary costs:** As explained in Chapter 6, nonsalary costs (including training and coaching, travel, and equipment/supplies) in the nonstudy setting are assumed to be identical to the costs of these resources in the study setting. However, it was also assumed that these resources would be used over a longer period of time and that their corresponding costs should be annualized over this period.

  The formula to calculate an annualization adjustment factor is as follows (where r is the annual depreciation rate and n is the number of years over which the resource is used):[13]

$$A(r,n) = \frac{r(1+r)^n}{r(1+r)^n - 1}$$

[13]For more on annualizing resource costs, see Levin and McEwan (2001), pp. 67-70.

Specifically, it was assumed that the training received by an ERO teacher would remain useful for four years, after which additional training would be necessary. Similarly, it was assumed that the equipment and supplies would be used for four years before needing to be replaced.[14]

[14]Resources supporting ERO teacher training activities (those under cost categories "Training" and "Travel") were subject to a 5 percent depreciation rate, which resulted in annualization adjustment rates for program Years 1 and 2 of 0.23 and 0.28, respectively. Those devoted to equipment and supplies were depreciated at a 3.2 percent rate yielding adjustment rates for Years 1 and 2 of 0.22 and 0.27.

Appendix M

Poststudy Adolescent Literacy Programming in the ERO Schools: Methodology and Additional Findings

This appendix discusses the methodology used to collect information about the literacy services offered in the study schools following the required two-year implementation of the Enhanced Reading Opportunities (ERO) programs. This is followed by a presentation of interview findings not discussed in Chapter 6, related to the following topics:

- The decision-making process that led to the continuation (or discontinuation) of the ERO programs
- The types of literacy programs being used at the study schools that discontinued the ERO programs
- The types of literacy-focused professional development opportunities available to teachers in the study districts after the end of the study
- Sources of funding for literacy programs after the conclusion of the study

Methodology

During the spring and early summer of 2009, the study team interviewed staff in ERO districts and schools to determine the status of adolescent literacy programming in the study schools after the conclusion of the grant-required ERO implementation period. District-level staff were interviewed to obtain a broader perspective on current adolescent literacy programming in the school district, while school-level staff were interviewed to provide specifics on the literacy services now offered at their school.

Interview protocols

The study team developed semi-structured protocols to guide the interviews. Two protocols were developed, one for district representatives and one for school-level staff (copies of the two protocols are included at the end of this appendix). Questions on the protocols focused on literacy activities during the school years since the ERO study-required implementation period, that is, school years 2007-2008 and 2008-2009. The protocols were designed to gather information about two primary constructs:

1. The continuation or discontinuation of the ERO intervention at the district and school level
2. The current status of programming for struggling adolescent readers in the district or schools

The protocols also sought information on the following topics:

- The process for deciding whether to continue the ERO program

- Modifications made to the Reading Apprenticeship Academic Literacy (RAAL) or Xtreme Reading intervention, if schools continued to use the program
- The literacy program that replaced the ERO program, if some other form of adolescent literacy services is now offered
- Funding sources for continuing the ERO intervention or its replacement programming

Sample and Analysis

Contact information gathered earlier in the study helped study team members to locate individuals who would be familiar with the ERO study and adolescent literacy programming since the end of the study. These contacts were initiated late in the 2008-2009 school year; interviews were then conducted during late spring and early summer 2009. Specifically:

- **District-level interviews:** A least one district-level interview was conducted in each of 10 study districts.[1] All interviewees had been in the district at the time of the study and were familiar with the Smaller Learning Communities (SLC) grant and the accompanying ERO study. District interviewees were either project coordinators for the SLC grants or secondary school English language arts or reading coordinators familiar with adolescent literacy services for struggling readers. All interviewees responded affirmatively about their familiarity with adolescent literacy programming in the district and with the ERO study specifically. At least one interviewee in each district had worked at the district level during the ERO demonstration years.
- **School-level interviews:** School-level interviews were conducted in 30 of the 34 ERO study schools.[2] Interviewees at the school level were principals, assistant principals, curriculum coordinators, or teachers familiar with the

[1]Multiple district-level interviews were conducted in seven of the districts to ensure complete data gathering. Nineteen (19) interviews were conducted in total across all study districts.

[2]The study team was unable to schedule interviews at three of the schools because summer breaks had started in those districts. There is no information on the fourth school because the former ERO teacher from this school had moved to a different school in the same district at the start of the 2007-2008 school year. Although familiar with the district's adolescent literacy programming in general, she could not report on continuation of the ERO program in her previous school. A total of 39 school-level interviews were conducted; eight of these interviews included two school representatives.

ERO study. All stated that they were familiar with the ERO study as it was implemented in their schools.

Two-person teams conducted the telephone interviews, which were audio-recorded with the permission of the interviewees. The interviewers were familiar with the study and had been trained to use the interview protocols.

Study team members then analyzed written interview notes and tape recordings (as needed) and consolidated interviewee responses according to the constructs underlying the interview protocols. Representative excerpts from the interviews were sorted by construct, after which the study team created data tables by construct that could be disaggregated by ERO program or by respondent.

Additional Findings

ERO Continuation Decisions

Interviews with district-level staff provided information on the process for deciding whether or not to continue using the ERO program (that is, whether the decision was made at the district level, the school level, or a combination of the district and school together). Findings by district are presented in Table M.1.

As seen in this table, the decision to continue/discontinue the ERO program was most often made by districts (four districts) or by individual schools (four districts). In the remaining two districts, the decision was made by the district and school together.

- The ERO programs' continued use was more likely in school districts where the decision about the continuation of the ERO programs was left to individual schools (as opposed to when school districts were also involved in this decision).

As seen in Table M.1, in all six districts where the decision to continue/discontinue the ERO programs involved district administrators (whether alone or together with the school), the ERO program was discontinued in at least one study school. In contrast, in the four districts where this decision was left to schools alone, the ERO program was discontinued in one school.

Adolescent Literacy Programming in Schools No Longer Implementing the ERO Programs

As reported in Chapter 6, formal use of the ERO programs was discontinued in 16 of the 30 schools for whom school-level interview data are available. At 14 of these schools,

The Enhanced Reading Opportunities Study

Appendix Table M.1

Summary of ERO Continuation Plans, by District

District	RAAL Programming	Xtreme Reading Programming	Continuation Decision
1	Discontinued	Discontinued	District
2	Discontinued	Discontinued	District
3	Continued (1) Discontinued (1)	Discontinued	District[a]
4	Discontinued	Continued (1) Discontinued (1)	District and school together
5	Discontinued	Discontinued	District and school together
6[b]	Continued	Discontinued[c]	District
7	Continued	Continued	School
8	Continued	Continued	School
9	Continued	Continued	School
10	Continued	Continued (1) Discontinued (1)	School

SOURCES: Interviews conducted with representatives of ERO districts and schools in spring and summer 2009.

NOTES:
[a]The decision to continue RAAL in one school was made at the school level.
[b]Programming was moved from ninth grade down to the junior high level across this district starting in 2008-2009.
[c]Program was discontinued after the 2007-2008 school year.

interviewees also described the replacement programs and approaches for addressing the needs of struggling adolescent readers during the poststudy period. These findings can be summarized as follows:

- In five schools (36 percent of interviewed noncontinuing schools), the ERO programs had been replaced with widely available commercial programs.

- Schools also opted for the following (less common) approaches: using a reading curriculum created by the district, offering reading support to 11th- and 12th-grade students who had not passed their state's high stakes assessment, and no longer offering a reading intervention for high school students. Also, in eight schools (57 percent of interviewed noncontinuing schools), teachers continued using what they referred to as "ERO strategies," without using the actual program.

Professional Development and Funding in the Poststudy Period

Beyond providing information on the continuation (or discontinuation) of the ERO programs, the interviews were also used to obtain a broader perspective on adolescent literacy programming at the ERO schools after the end of the study. This section, in particular, discusses the extent to which districts and schools remained committed to helping struggling adolescent readers more broadly, as measured by professional development opportunities provided to teachers as well as the funding of literacy interventions.

Professional Development Opportunities

Interviews with school staff also provided information on the extent to which literacy-focused professional development opportunities were available to teachers after the end of the study. This provides an indication of schools' commitment to helping struggling adolescent readers:

- At 22 high schools (73 percent of interviewed schools), formal literacy-related professional development was offered to teachers after the demonstration ended, with 11 schools specifically offering training focused at strategies for embedding literacy in the content areas.[3]

Information was also available on the extent to which schools continued to build their internal capacity for literacy instruction, as reflected by coaching and mentoring:

- Teachers in 22 of the schools (73 percent of interviewed schools) mentioned that they had continued to have access to some form of coaching.[4]

---

[3]In particular, interviewees at three schools that had replaced the ERO program said they had been trained for the new commercial intervention programs they were teaching.

[4]For a subset of these schools, additional information was available on the type of coach that was used. In nine schools, interviewees reported that they were supported by a school-based coach; four schools used a

(continued)

- Four former ERO teachers said that they provided training or mentoring to other teachers in their schools or district.

Funding Sources for Adolescent Literacy Programming

During the study-required implementation period, the implementation of the ERO programs was explicitly supported by grant funds. After the end of this period, all district-level staff asserted that their districts were continuing to focus efforts toward addressing the needs of adolescent struggling readers. However, in seven of the 10 study districts, funding challenges were cited as a consideration in decisions made about adolescent literacy programming.

Interviewees in nine district offices were able to provide additional information on sources of funding for their adolescent literacy initiatives since the end of the study. The funding sources mentioned in these interviews are the following:

- Remaining funds from the SLC grant (5 districts)[5]
- Districts' general budget (7 districts)
- Title I and II funds (5 districts)
- Grants from sources other than the SLC (5 districts)

district-based coach; and three schools reported that they were supported by a coach associated with a new commercial reading intervention program.

[5]Of these five districts, four have continued to offer the ERO program.

Enhanced Reading Opportunities Study
District Coordinator Follow-Up Interview

Thank you for taking the time to speak with me today. I am [interviewer name]. As you may know, staff from the American Institutes for Research (AIR) and MDRC visited your district in 2005-06 and 2006-07 to monitor the implementation of the ERO courses that were used in [name schools] as part of the federally-funded Smaller Learning Communities (SLC) grant. It has been two years since we were in your district, so we wanted to follow up on what is being done to address issues of adolescent literacy and also the status of the ERO course in your district and within each ERO school.

Are you a good person to talk to about these issues? If not, who would be able to help me?

[*If no, thank the person for his/her time and end interview.]*

FOR ALL INTERVIEWEES:

Before we continue, if it's ok with you, we would like to audio-record our interview today. We will use the recording as a back up to help us clean our notes and fill in anything I missed. The recording will not be shared with anyone outside of the project staff. Is this ok?

If ok, say: I am speaking with [interviewee name] from [school/district name] on [date].

Highlight: YES NO

Our first set of questions is specifically about adolescent literacy initiatives in general in your district.

Questions	Notes
1. Are you familiar with any initiatives in your district to address the needs of adolescents who struggle with reading and writing? If YES, continue. If NO, ask who else might be able to answer questions.	
2. Please briefly describe some of the activities that characterize the district's efforts related to adolescent literacy and tell us the goals of these efforts. (*Possible responses*: interventions and other help for struggling readers; PD for all teachers on the literacy needs of students and differentiated instruction; PD on reading-in-the-content areas; use of literacy coaches in schools, etc.) o If one or more interventions are in use, what is/are they? o Who [all] is providing the training, PD, and support for teachers? o What funding sources are used for these initiatives? o Are there literacy or academic coaches and how they are funded? (*Possible responses*: SLC grant funds, district resources, school resources, other grants)	
3. In what ways are the adolescent literary efforts aligned with other initiatives in the district (e.g., in math and science)? Is there general buy-in for the efforts, including from content area teachers?	
4. From your perspective, how successful have these efforts been in improving students' literacy skills? What efforts, if any, are being used to monitor these initiatives?	

The next set of questions relate specifically to the Enhanced Reading Opportunities (ERO) study that was conducted in your district. The implementation phase of the study ended with the 2006 – 2007 school year. We are interested in knowing about use of the programs – Xtreme Reading and Reading Apprenticeship Academic Literacy (RAAL) – in the following two school years.

5. To the best of your knowledge, did any of the ERO schools continue using Xtreme Reading or RAAL? Did any switch their program and adopt the one used in the other ERO schools? • How were decisions made to [continue or discontinue] use? IF YES THEY CONTINUED USING ONE OR BOTH, ask the following questions. IF NO, skip to 10.] *Notes should clearly indicate any differences in implementation for school years 2006 – 2007 and 2007 – 2008.*	
6. What can you tell us about *how* the programs were or are being implemented? PROMPTS: What's their current structure/format? What are their goals? Are schools using the program(s) in the same way they were used in the study? What sorts of changes have been made to schedules, number of students served, or the instruction? Have schools added additional teachers or sections of the course? In what ways have teachers been trained and supported? *Notes should clearly indicate any differences in implementation for school years 2006 – 2007 and 2007 – 2008.*	

7. What challenges have there been in continuing to implement the ERO interventions? What has made it easier to continue to implement ERO? (*Possible responses*: difficulty finding appropriate teachers, students resist enrollment in a reading course; scheduling; training teachers; maintaining equipment and having enough supplies)	
8. What steps are in place in the district to monitor the success of these programs? As far as you know, will the programs be used again next year, 2009 - 2010?	

This information is very helpful. We have one final area we'd like to discuss: <u>what you experienced or remember about participating in the ERO study</u>. During the implementation years, students were assigned at random to either receive the reading interventions or to take their regular elective courses.

9. What were the challenges and benefits to participating in this study?	
10. What advice would you offer other districts considering participation in a random assignment study? What advice would you offer to researchers conducting random assignment studies about how to work with schools and districts?	
11. Would you want your district to be involved in future studies like this? What would make such an opportunity more or less appealing?	

Thank you for sharing this information with us. We are almost finished.

12. Is there anything else you'd like to tell us about this district, the participating schools, or the ERO study that we haven't already covered?	
13. Is there anyone else we should talk to about literacy initiatives in the district or about the ERO study?	

Enhanced Reading Opportunities Study

School Follow-Up Interview

Thank you for taking the time to speak with me today. I am [interviewer name]. As you may remember, staff from the American Institutes for Research (AIR) and MDRC visited your school in 2005-06 and 2006-07 to monitor the implementation of the ERO courses that you taught. It has been two years since we were in your school, so we wanted to follow up on what is being done to address issues of adolescent literacy and also the status of the ERO course in your school.

Are you a good person to talk to about adolescent literacy in your school? If NO, who would be able to help me?

FOR ALL INTERVIEWEES:

Before we continue, if it's ok with you, we would like to audio-record our interview today. We will use the recording as a back up to help us clean our notes and fill in anything I missed. The recording will not be shared with anyone outside of the project staff. Is this ok?

If ok, say: I am speaking with [interviewee name] from [school/district name] on [date].

Highlight: YES NO

Our first set of questions is specifically about adolescent literacy initiatives in general at your school.

Questions	Notes
1. Are you familiar with any initiatives in the school to address the needs of adolescents who struggle with reading and writing? If YES, continue. If NO, ask who else might be able to answer questions.	
2. Please tell us about any initiatives in place in your school to address the needs of adolescents who struggle with reading and writing. • Are there specific intervention programs in use for struggling readers? What are they? • What has motivated the school to implement these initiatives? • How are these initiatives monitored?	
3. From your perspective, in what ways are the adolescent literary efforts aligned with other initiatives in your school (e.g., in math or science)? Is there general buy-in for the efforts, especially from content area teachers?	
4. During 2005 – 2006 and 2006 – 2007, your school implemented [name of program] as part of the ERO study. We're interested in knowing what happened after implementation as part of the study ended, that is, in school years 2007 – 2008 and 2008 – 2008. Did the school continue implementing the ERO program in either or both school years? If YES, continue with these sub-questions and *be sure to get as much differentiation as possible about each year*.	

If NO, skip to questions 5. • In what grades is it being taught? • Who is teaching the course? • Have any additional teachers been added to teach the course? • How is the program being paid for? • How many sections are there and how many students are served? • What other courses, if any, does the ERO teacher or do the ERO teachers teach? • Does the ERO teacher or do the ERO teachers have contact with the ERO developers? • How was the decision to [continue or discontinue] the program made at your school? • In what ways, if any, has the course been modified now that the ERO study has ended?	

5. Now two years after the ERO study ended, what suggestions do you have to change and improve the reading intervention offered at your school?	
6. What is your school planning to do to help struggling readers in the next school year, 2008 - 2009?	

This information is very helpful. We have one final area we'd like to discuss: what you experienced or remember about participating in the ERO study. During the implementation years, students were assigned at random to either receive the reading interventions or to take their regular elective courses.

7. What were the challenges and benefits to participating in this study?	
8. What advice would you offer other schools considering participation in a random assignment study? What advice would you offer to researchers conducting random assignment studies about how to work with schools and districts?	
9. Would you want to participate in a similar study at your school in the future? What would make such an opportunity more or less appealing?	

Thank you for sharing this information with us. We are almost finished.

10. Is there anything else you'd like to tell us about this school, its literacy initiatives, or the ERO study that we haven't already covered?	
11. Is there anyone else we should talk to about literacy initiatives in the school or about the ERO study?	