

Appendix A

ERO Student Follow-Up Survey Measures

Two surveys were administered during the second year of the Enhanced Reading Opportunities (ERO) study. The Student Background Questionnaire, completed by all the student participants early in the 2006-2007 school year, included questions to ensure that random assignment was effective in dividing students evenly between the ERO and non-ERO groups.

This appendix describes the development of measures created from the second student-level survey, the ERO Student Follow-Up Questionnaire. This survey was administered to students in the second cohort of the study near the end of their ninth-grade year during the spring of 2007. The questions on this survey were intended to assess whether students participated in literacy support activities during the school year and to measure student attitudes and behaviors related to reading activities. A variety of measures were constructed by combining conceptually and empirically linked items from the survey. The ERO study team used a three-step process for defining and constructing the measures discussed in this appendix:

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

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

Measures of Self-Reported Participation in Supplemental Literacy Support Activities

This section of the appendix describes four measures which assess the duration and frequency of student participation in supplemental literacy support activities: (1) attending a reading or writing class that took place in school, (2) working with a reading or writing tutor in school, (3) attending a reading or writing class that took place outside school, and (4) working with a reading or writing tutor outside school. Questions about the first of these activities were intended to determine whether students identified themselves as being enrolled in the ERO classes or similar types of classes that may have been offered in their high schools. Student reports about their participation in the other three activities were intended to provide an indication of the extent to which they utilized supplemental literacy support activities outside the ERO classes or similar classes that may have been offered in the participating high schools. The overall contrast between the ERO and non-ERO groups on these measures provides an indication of whether the ERO programs added literacy support activities to the landscape of what

would have been available to students without the programs, at least as reported by the students in the study sample.

Each of the four measures was created based on three survey items. The first item (questions 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 A.1 lists the participation values calculated for every combination of answers to the questions about duration and frequency. The columns represent duration, “how long” a student received extra help (questions 6, 9, 12, and 15). The rows represent frequency, “how often” a student received that help (questions 7, 10, 13, and 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.

The Enhanced Reading Opportunities Study

Appendix Table A.1

Intensity Values for Supplemental Literacy Support Measures

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

Measures of Self-Reported Reading Behaviors

The ERO Student Follow-Up Questionnaire 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 for 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. It also focused on groups of items for which student responses were highly correlated (that is, groups of items that were correlated with Cronbach's $\alpha \geq .70$). The seven items used to construct a measure of in-school reading frequency were correlated with Cronbach's $\alpha = .71$, and the seven items used to construct a measure of out-of-school reading were correlated with Cronbach's $\alpha = .75$.

The study team also developed a measure of the frequency with which a student used reading strategies in reading for other courses. The six strategies included in the measure are

often used by proficient readers and are strategies that 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 = .71)

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 that 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 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 = .75)

This construct is designed to measure the frequency with which students read extended texts outside school. It combines student responses to questions about how often they read seven types of text during the previous month. Each possible answer is converted into a value based on the approximate number of sessions the student reported reading a given type of ma-

¹Biancarosa and Snow (2004).

terial 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 past month, about how OFTEN did you READ each of the following when you were not in school and not doing homework?

- b. Fiction books or stories
- c. 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 = .77)

This construct attempts to measure the degree to which students use reading strategies in which they reflect on what they are reading and ask questions of the text to better understand what they read. These measures both are consistent with the strategies taught by the ERO programs and are seen as antecedents to reading proficiency. The two questions that make up this measure were asked in the context of the reading that students do for their English class and for the reading they do for one other core-content-area class (history, science, or math), for a total of four items.

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.
- e. 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"

Question 18. For which one of the following classes did you do the most reading during the past school year?

4. History (or Social Studies)
5. Science
6. 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.
- e. 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"

Other Measures on Student Perceptions About Reading

The study team developed two other measures to assess the impact of the ERO program on students' perceptions of reading. The creation of each of these measures is described below.

Reading to Learn (5 items, Cronbach's Alpha = .80)

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

Question 2. Please indicate how much you DISAGREE or AGREE with the statements below about reading and writing.

- a. When I read books, I learn a lot.
- f. I read because it helps me do better in my classes.
- g. I read to see what is going on in the world, the country, and/or my community.
- i. I read in order to learn new things.
- j. I read to learn how other people see things.

Scale: 1 = "Strongly Disagree" to 4 = "Strongly Agree"

Reading to Enjoy (2 items, Cronbach's Alpha = .82)

This construct was designed to measure whether or not a student enjoys reading. It was created by averaging student responses to the items below. If a student did not respond to at least one of the items, the measure was coded as missing.

Question 2. Please indicate how much you DISAGREE or AGREE with the statements below about reading and writing.

- b. Reading is one of my favorite activities.
- h. I read because I enjoy it.



**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.

- graduate from high school
- vocational or technical training (e.g. electrician, hairdresser, chef, pre-school teacher)
- some college
- graduate from a business or two-year college
- graduate from a four-year college
- get a master's degree
- 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. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| b. Reading is one of my favorite activities. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| c. Writing things like stories or letters is one of my favorite activities. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| d. Writing helps me share my ideas. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| e. When I have free time, I rarely choose to read over doing other activities. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| f. I read because it helps me do better in my classes. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| g. I read to see what is going on in the world, the country, and/or my community. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| h. I read because I enjoy it. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| i. I read in order to learn new things. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| j. I read to learn how other people see things. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

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 ○ |

| | Yes | No |
|---|--|---|
| (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? | <input type="radio"/> If YES, please continue to question 6 | <input type="radio"/> If NO, please continue to question 8 |

| | One month or less | A couple of months | One semester or term | Most of the year | All year |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| (6) For how LONG did you get this help with reading and writing? | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

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

| | Yes | No |
|---|--|--|
| (8) Did an adult <u>in your school</u> help you individually with your reading and writing this year, like a tutor? | <input type="radio"/> If YES, please continue to question 9 | <input type="radio"/> If NO, please continue to question 11 |

| | One month or less | A couple of months | One semester or term | Most of the year | All year |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| (9) For how LONG did you get this help with reading and writing? | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

| | Less than once a month | Once a month | Every other week | Once a week | Twice a week | 3-4 times a week | Every day |
|--|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| (10) How OFTEN did you get this help with reading and writing? | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

| | Yes | No |
|---|--|---------------------------------------|
| (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? | 1 <input type="radio"/> | 2 <input type="radio"/> |
| | If YES, please continue to question 12 | If NO, please continue to question 14 |

| | One month or less | A couple of months | One semester or term | Most of the year | All year |
|---|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| (12) For how LONG did you get this help with reading and writing? | 1 <input type="radio"/> | 2 <input type="radio"/> | 3 <input type="radio"/> | 4 <input type="radio"/> | 5 <input type="radio"/> |

| | Less than once a month | Once a month | Every other week | Once a week | Twice a week | 3-4 times a week | Every day |
|--|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| (13) How OFTEN did you get this help with reading and writing? | 1 <input type="radio"/> | 2 <input type="radio"/> | 3 <input type="radio"/> | 4 <input type="radio"/> | 5 <input type="radio"/> | 6 <input type="radio"/> | 7 <input type="radio"/> |

| | Yes | No |
|---|--|---------------------------------------|
| (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? | 1 <input type="radio"/> | 2 <input type="radio"/> |
| | If YES, please continue to question 15 | If NO, please continue to question 17 |

| | One month or less | A couple of months | One semester or term | Most of the year | All year |
|---|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| (15) For how LONG did you get this help with reading and writing? | 1 <input type="radio"/> | 2 <input type="radio"/> | 3 <input type="radio"/> | 4 <input type="radio"/> | 5 <input type="radio"/> |

| | Less than once a month | Once a month | Every other week | Once a week | Twice a week | 3-4 times a week | Every day |
|--|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| (16) How OFTEN did you get this help with reading and writing? | 1 <input type="radio"/> | 2 <input type="radio"/> | 3 <input type="radio"/> | 4 <input type="radio"/> | 5 <input type="radio"/> | 6 <input type="radio"/> | 7 <input type="radio"/> |

This section is about your classes in school this year.

(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 <u>English</u> class. | 1 <input type="radio"/> | 2 <input type="radio"/> | 3 <input type="radio"/> | 4 <input type="radio"/> |
| b. When I don't understand a word while reading for <u>English</u> class, I try to break the word down into smaller pieces. | 1 <input type="radio"/> | 2 <input type="radio"/> | 3 <input type="radio"/> | 4 <input type="radio"/> |
| c. To help me understand what I'm reading for <u>English</u> class, I try to connect the things that are unfamiliar to me with things I already know. | 1 <input type="radio"/> | 2 <input type="radio"/> | 3 <input type="radio"/> | 4 <input type="radio"/> |
| d. While reading for <u>English</u> class, I rarely make predictions about what will come next in a passage. | 1 <input type="radio"/> | 2 <input type="radio"/> | 3 <input type="radio"/> | 4 <input type="radio"/> |
| e. When I'm reading for <u>English</u> class I stop once in a while and go over what I have read. | 1 <input type="radio"/> | 2 <input type="radio"/> | 3 <input type="radio"/> | 4 <input type="radio"/> |
| f. When I don't know the meaning of a word while reading for <u>English</u> class, I often look at other words in the sentence or paragraph to help me understand. | 1 <input type="radio"/> | 2 <input type="radio"/> | 3 <input type="radio"/> | 4 <input type="radio"/> |
| g. I try to identify whether what I'm reading for <u>English</u> class is fact or opinion. | 1 <input type="radio"/> | 2 <input type="radio"/> | 3 <input type="radio"/> | 4 <input type="radio"/> |
| h. I rarely stop to summarize a passage or paragraph while reading for <u>English</u> class. | 1 <input type="radio"/> | 2 <input type="radio"/> | 3 <input type="radio"/> | 4 <input type="radio"/> |

(18) For which one of the following classes did you do the most reading during the past school year?

Mark (✓) one answer.

History (or Social Studies)

Science

Math

(19) Please indicate how much you DISAGREE or AGREE with the following statements about the class you chose in Question 18. 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. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| b. When I don't understand a word while reading for class, I try to break the word down into smaller pieces. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 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. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| d. While reading for class, I rarely make predictions about what will come next in a passage. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| e. When I'm reading for class I stop once in a while and go over what I have read. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 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. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| g. I try to identify whether what I'm reading for class is fact or opinion. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| h. I rarely stop to summarize a passage or paragraph while reading for class. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

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

(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. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| b. Compared to work I do for other subjects at school, I find the work I do for ERO to be interesting. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| c. Compared with what I learn in my other subjects at school, I find what I learn in ERO to be useful. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

THANK YOU!!!

Appendix B

Follow-Up Test and Survey Response Analysis

The two main data sources for the second-year impact analysis of the Enhanced Reading Opportunities (ERO) study are the Group Reading Assessment and Diagnostic Examination (GRADE) assessment of student reading skills and the student follow-up survey. Both the test and the survey were administered late in the 2006-2007 school year. Overall, 81 percent of the full study sample completed the test and survey, including 83 percent of students in the ERO program group and 79 percent of students in the non-ERO group. The lack of a 100 percent response rate combined with the discrepancy between response rates for the ERO and non-ERO student groups raises two concerns: Are the respondents representative of the full study sample? Are there systematic pre-program differences between respondents in the ERO and non-ERO groups?

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

Follow-Up Test and Survey Response Rates

Efforts were made to collect both test and survey data from all 2,679 students who make up the full study sample — ninth-grade students who consented to be in the ERO program and had pretest reading comprehension scores between the fourth- and seventh-grade levels. Sections of 25 to 30 students from both the ERO and the non-ERO group were tested and surveyed together in their high schools. The test and survey administrations took place during the school day and were proctored by members of the ERO study team. The ERO study team spent up to four days at each school locating, testing, and surveying students who did not attend the originally scheduled session.

In all, 2,160 students (81 percent of the full study sample) completed both the follow-up test and the survey. An additional 11 students completed only the follow-up test, and one student completed only the survey. Due to the similarity in response rates for the follow-up test and the survey, the nonresponse analysis in this appendix focuses on the response rate for the test. Results for the survey response and the combined response are virtually the same.

Appendix Table B.1 shows the follow-up test response rates for all 34 participating high schools combined and for the groups of schools using Reading Apprenticeship Academic Literacy (RAAL) and Xtreme Reading, respectively. Overall, 83 percent of students in the ERO group took the follow-up test, compared with 79 percent of students in the non-ERO group. The 4 percentage point difference is statistically significant (the p-value is less than or equal to 5 percent). The RAAL and Xtreme Reading schools had similar response rates for their ERO and

The Enhanced Reading Opportunities Study

Appendix Table B.1

**Response Rates of Students in Cohort 2
Full Study Sample**

| | ERO Group | Non-ERO Group | Difference | P-Value for the Difference |
|--|-----------|---------------|------------|----------------------------|
| All schools | 82.7 | 79.1 | 3.6 * | 0.018 |
| Reading Apprenticeship schools | 82.4 | 79.4 | 3.0 | 0.163 |
| Xtreme Reading schools | 83.0 | 78.7 | 4.3 * | 0.049 |
| Overage for grade ^a | 73.4 | 71.0 | 2.4 | 0.455 |
| Not overage for grade | 86.4 | 82.8 | 3.6 * | 0.031 |
| Language other than English spoken at home | 84.3 | 81.5 | 2.8 | 0.190 |
| English only spoken at home | 81.2 | 76.8 | 4.4 * | 0.047 |
| Baseline reading comprehension score | | | | |
| 2.0-3.0 years below grade level | 85.2 | 81.3 | 3.9 | 0.131 |
| 3.1-4.0 years below grade level | 83.7 | 82.9 | 0.8 | 0.767 |
| 4.1-5.0 years below grade level | 80.1 | 73.9 | 6.1 * | 0.016 |
| Teachers having taught two full years of ERO | 83.6 | 80.9 | 2.7 | 0.115 |
| Teachers having taught less than two full years of ERO | 80.1 | 74.1 | 6.0 | 0.054 |
| Stronger implementation schools ^b | 80.4 | 75.8 | 4.6 | 0.070 |
| Weaker implementation schools ^c | 84.0 | 81.1 | 3.0 | 0.116 |
| Sample size | 1,529 | 1,150 | | |

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

NOTES: This table represents the response rates for the follow-up GRADE assessment which was administered in spring 2007 at the end of students' ninth-grade year. The follow-up student questionnaire was also administered at that time. The difference in response rates between the test and survey is negligible.

The estimated differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. 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 (i.e., schools) as the basis for the adjustment.

Rounding may cause slight discrepancies in calculating sums and differences.

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

^bThe ERO programs in these schools were deemed to have reached an implementation level that was very well aligned to both the classroom learning environment and comprehension instruction dimensions of the program model by the spring site visit, and their ERO program began operating within 2 weeks of the start of the school year.

^cThe implementation fidelity of the ERO programs in these schools was deemed to be less than very well aligned to the classroom learning environment and/or comprehension instruction dimensions of the program model by the spring site visit, and/or their ERO program began operating 2 weeks or more after the start of the school year.

non-ERO group students. The difference in response rates between the ERO and non-ERO groups is statistically significant for the Xtreme Reading schools but not for the RAAL schools.

The primary reason that students did not complete the follow-up test or survey is that they were no longer enrolled in a high school participating in the ERO study.¹ In all, 20 percent of the students in the study sample were no longer enrolled in an ERO high school at the time of the follow-up test and survey administrations. These rates are similar for the ERO group (21 percent) and the non-ERO group (20 percent). Of the students who were no longer enrolled in an ERO school, only 28 percent completed the follow-up test (compared with 95 percent of those who remained enrolled in an ERO school). As in the full sample, completion rates among students who were no longer enrolled in an ERO school differ for the ERO group (30 percent) and the non-ERO group (26 percent). Also like the full sample findings, this difference in response rates is concentrated in the Xtreme Reading sites, where 30 percent of the ERO group completed the follow-up test, compared with 24 percent of the non-ERO group. Among students who remained enrolled in an ERO school, response rates also differ between the two treatment groups: 96 percent for the ERO group and 92 percent for the non-ERO group. Unlike the full sample findings, however, the difference in response rates between treatment groups among students who remained enrolled in an ERO high school is approximately the same for Xtreme Reading and RAAL schools.

One factor that may influence the interpretation of the impact findings presented in this report is whether students who completed the follow-up test and survey are representative of the full study sample. This question was addressed in two ways. First, respondents and nonrespondents were compared directly on a range of background characteristics. The results for the full study sample are shown in Appendix Table B.2. Overall, the table indicates that nonrespondents are more likely than respondents to have characteristics associated with a risk of school failure. For example, a higher percentage of nonrespondents are overage for the ninth grade (41 percent, compared with 26 percent for respondents), thus indicating that nonrespondents are more likely to have been retained in a prior grade. In addition, a smaller percentage of nonrespondents have a mother who completed some postsecondary education (27 percent, compared with 32 percent for respondents) or a father who completed some postsecondary education (17 percent, compared with 22 percent for respondents). On average, nonrespondents also had lower reading comprehension test scores at baseline than students who completed the follow-up test (84 points for nonrespondents and 85 points for respondents). All of these differences are statistically significant at the 5 percent level. Appendix Tables B.3 and B.4 compare the respondents and nonrespondents in RAAL schools and Xtreme Reading schools, respectively.

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

The Enhanced Reading Opportunities Study

Appendix Table B.2

Characteristics of Students in Cohort 2:
Differences Between Respondents and Nonrespondents

| Characteristic | Respondents | Non- Respondents | Difference | P-Value for the Difference |
|--|-------------|---------------------|------------|-------------------------------|
| Race/ethnicity (%) | | | | |
| Hispanic | 31.6 | 27.2 | 4.4 * | 0.006 |
| Black, non-Hispanic | 46.6 | 48.7 | -2.1 | 0.256 |
| White, non-Hispanic | 15.2 | 16.8 | -1.5 | 0.325 |
| Other | 6.6 | 7.3 | -0.7 | 0.545 |
| Gender (%) | | | | |
| Male | 51.3 | 49.2 | 2.1 | 0.401 |
| Female | 48.7 | 50.8 | -2.1 | 0.401 |
| Average age (years) | 14.7 | 14.9 | -0.2 * | 0.000 |
| Overage for grade ^a (%) | 26.2 | 40.6 | -14.4 * | 0.000 |
| Language other than English spoken at home (%) | 49.7 | 48.1 | 1.6 | 0.479 |
| Language spoken at home missing (%) | 1.2 | 5.3 | -4.2 * | 0.000 |
| Mother's education level (%) | | | | |
| Did not finish high school | 17.4 | 20.9 | -3.5 | 0.066 |
| High school diploma or GED certificate | 24.6 | 23.3 | 1.3 | 0.535 |
| Completed some postsecondary education | 32.3 | 27.3 | 5.0 * | 0.028 |
| Don't know | 23.9 | 22.8 | 1.1 | 0.597 |
| Missing | 1.8 | 5.8 | -4.0 * | 0.000 |
| Father's education level (%) | | | | |
| Did not finish high school | 16.2 | 17.5 | -1.3 | 0.460 |
| High school diploma or GED certificate | 22.2 | 20.6 | 1.5 | 0.459 |
| Completed some postsecondary education | 21.6 | 16.6 | 5.1 * | 0.011 |
| Don't know | 37.5 | 39.3 | -1.7 | 0.477 |
| Missing | 2.5 | 6.0 | -3.5 * | 0.000 |
| GRADE reading comprehension ^b | | | | |
| Average standard score | 84.7 | 83.6 | 1.1 * | 0.000 |
| <i>Corresponding grade equivalent</i> | <i>4.9</i> | <i>4.7</i> | | |
| <i>Corresponding percentile</i> | <i>14</i> | <i>13</i> | | |
| 2.0 - 3.0 years below grade level (%) | 32.5 | 28.2 | 4.2 | 0.065 |
| 3.1 - 4.0 years below grade level (%) | 28.5 | 24.9 | 3.6 | 0.105 |
| 4.1 - 5.0 years below grade level (%) | 39.0 | 46.9 | -7.8 * | 0.001 |
| Sample size | 2,171 | 508 | | |

(continued)

Appendix Table B.2 (continued)

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

NOTES: Baseline data for students in 30 of 34 high schools were collected in spring 2006 (during students' eighth-grade year); baseline data for students in the remaining four schools were collected in fall 2006 (at the start of students' ninth-grade year).

The estimated differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The values in the column labeled "Respondents" are the observed means for students in the study sample with follow-up data on the GRADE assessment (respondents). The "Non-Respondents" values in the next column are the regression-adjusted means for students in the study sample without follow-up data on the GRADE assessment (non-respondents), using the observed distribution of respondents across random assignment blocks (i.e., schools) as the basis for the adjustment.

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

Rounding may cause slight discrepancies in calculating sums and differences.

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

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

A second and more comprehensive strategy for assessing differences between respondents and nonrespondents is to use multiple regression to determine the extent to which the average characteristics of students who completed the follow-up test differ systematically from those of students who did not. This analysis was carried out for the full group of schools in the study and separately for the schools using RAAL and Xtreme Reading, respectively. The results are presented in Appendix Table B.5. It indicates that response rates differ by some background characteristics, including age, race/ethnicity, and baseline test scores. More important, the overall F-test for each regression indicates that there are systematic differences between the respondents and nonrespondents.

In summary, the response analysis indicates that students who completed the follow-up test and survey are not fully representative of the full study sample of 2,679 students. Thus, some caution should be exercised when attempting to generalize the findings beyond those sample members who are included in the impact analysis. Nevertheless, the overall response rates show that follow-up data are available for 81 percent of the students in the study sample, making the results reflective of the behavior of most of the targeted students.

Appendix F presents an assessment of the sensitivity of the impact findings to differences between students who completed the follow-up test and those who did not. The appendix presents estimated impacts that are weighted for differential response rates by high school,

The Enhanced Reading Opportunities Study

Appendix Table B.3

Characteristics of Students in Cohort 2:
Differences Between Respondents and Nonrespondents,
Reading Apprenticeship Schools

| Characteristic | Respondents | Non- Respondents | Difference | P-Value for the Difference |
|--|-------------|---------------------|------------|-------------------------------|
| Race/ethnicity (%) | | | | |
| Hispanic | 31.5 | 26.3 | 5.1 * | 0.023 |
| Black, non-Hispanic | 47.4 | 48.8 | -1.3 | 0.618 |
| White, non-Hispanic | 14.1 | 19.1 | -5.1 * | 0.018 |
| Other | 7.0 | 5.8 | 1.2 | 0.474 |
| Gender (%) | | | | |
| Male | 51.1 | 50.6 | 0.5 | 0.879 |
| Female | 48.9 | 49.4 | -0.5 | 0.879 |
| Average age (years) | 14.7 | 15.0 | -0.2 * | 0.000 |
| Overage for grade ^a (%) | 27.7 | 42.3 | -14.6 * | 0.000 |
| Language other than English spoken at home (%) | 48.4 | 47.6 | 0.8 | 0.785 |
| Language spoken at home missing (%) | 1.1 | 5.9 | -4.9 * | 0.000 |
| Mother's education level (%) | | | | |
| Did not finish high school | 18.1 | 25.2 | -7.1 * | 0.009 |
| High school diploma or GED certificate | 25.4 | 23.5 | 1.9 | 0.526 |
| Completed some postsecondary education | 30.4 | 24.8 | 5.6 | 0.074 |
| Don't know | 24.1 | 20.6 | 3.6 | 0.222 |
| Missing | 2.0 | 5.9 | -4.0 * | 0.001 |
| Father's education level (%) | | | | |
| Did not finish high school | 15.4 | 20.6 | -5.1 * | 0.042 |
| High school diploma or GED certificate | 23.8 | 18.9 | 4.9 | 0.095 |
| Completed some postsecondary education | 20.5 | 15.1 | 5.4 * | 0.047 |
| Don't know | 37.5 | 38.6 | -1.1 | 0.741 |
| Missing | 2.8 | 6.8 | -4.1 * | 0.002 |
| GRADE reading comprehension ^b | | | | |
| Average standard score | 84.8 | 83.4 | 1.4 * | 0.001 |
| <i>Corresponding grade equivalent</i> | <i>4.9</i> | <i>4.6</i> | | |
| <i>Corresponding percentile</i> | <i>15</i> | <i>12</i> | | |
| 2.0 - 3.0 years below grade level (%) | 33.8 | 29.1 | 4.7 | 0.146 |
| 3.1 - 4.0 years below grade level (%) | 27.1 | 22.8 | 4.3 | 0.161 |
| 4.1 - 5.0 years below grade level (%) | 39.1 | 48.0 | -8.9 * | 0.009 |
| Sample size | 1,115 | 262 | | |

(continued)

Appendix Table B.3 (continued)

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

NOTES: Baseline data for students in 30 of 34 high schools were collected in spring 2006 (during students' eighth-grade year); baseline data for students in the remaining four schools were collected in fall 2006 (at the start of students' ninth-grade year).

The estimated differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The values in the column labeled "Respondents" are the observed means for students in the study sample with follow-up data on the GRADE assessment (respondents). The "Non-Respondents" values in the next column are the regression-adjusted means for students in the study sample without follow-up data on the GRADE assessment (non-respondents), using the observed distribution of respondents across random assignment blocks (i.e., schools) as the basis for the adjustment.

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

Rounding may cause slight discrepancies in calculating sums and differences.

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

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

overage for grade, pretest scores, and research status. These analyses yield impact estimates that are similar to those presented in the text of the report.

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

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

When completion rates for follow-up data collection are less than 100 percent, a key question underlying the impact analyses is: Do the response rates preserve the random assignment design? In other words, does the sample of students who completed the follow-up test and survey exhibit the same lack of systematic differences between the ERO and non-ERO groups, both overall and for groups of sites using RAAL and Xtreme Reading? To address this question, multiple regression was used to assess whether there are systematic differences in background

The Enhanced Reading Opportunities Study

Appendix Table B.4

Characteristics of Students in Cohort 2:
Differences Between Respondents and Nonrespondents,
Xtreme Reading Schools

| Characteristic | Respondents | Non- Respondents | Difference | P-Value for the Difference |
|--|-------------|---------------------|------------|-------------------------------|
| Race/ethnicity (%) | | | | |
| Hispanic | 31.7 | 28.2 | 3.5 | 0.116 |
| Black, non-Hispanic | 45.6 | 48.6 | -2.9 | 0.259 |
| White, non-Hispanic | 16.5 | 14.3 | 2.2 | 0.320 |
| Other | 6.2 | 9.0 | -2.8 | 0.107 |
| Gender (%) | | | | |
| Male | 51.5 | 47.7 | 3.8 | 0.294 |
| Female | 48.5 | 52.3 | -3.8 | 0.294 |
| Average age (years) | 14.7 | 14.9 | -0.2 * | 0.000 |
| Overage for grade ^a (%) | 24.5 | 38.8 | -14.2 * | 0.000 |
| Language other than English spoken at home (%) | 51.0 | 48.7 | 2.3 | 0.461 |
| Language spoken at home missing (%) | 1.2 | 4.6 | -3.4 * | 0.001 |
| Mother's education level (%) | | | | |
| Did not finish high school | 16.7 | 16.3 | 0.4 | 0.875 |
| High school diploma or GED certificate | 23.8 | 23.1 | 0.7 | 0.816 |
| Completed some postsecondary education | 34.3 | 29.9 | 4.4 | 0.184 |
| Don't know | 23.7 | 25.2 | -1.5 | 0.625 |
| Missing | 1.6 | 5.6 | -4.0 * | 0.000 |
| Father's education level (%) | | | | |
| Did not finish high school | 17.0 | 14.3 | 2.7 | 0.308 |
| High school diploma or GED certificate | 20.5 | 22.5 | -2.0 | 0.493 |
| Completed some postsecondary education | 22.8 | 18.1 | 4.7 | 0.111 |
| Don't know | 37.6 | 40.0 | -2.4 | 0.498 |
| Missing | 2.2 | 5.2 | -3.0 * | 0.012 |
| GRADE reading comprehension ^b | | | | |
| Average standard score | 84.7 | 83.8 | 0.8 * | 0.042 |
| <i>Corresponding grade equivalent</i> | <i>4.9</i> | <i>4.7</i> | | |
| <i>Corresponding percentile</i> | <i>14</i> | <i>13</i> | | |
| 2.0 - 3.0 years below grade level (%) | 31.1 | 27.3 | 3.7 | 0.251 |
| 3.1 - 4.0 years below grade level (%) | 30.0 | 27.1 | 2.9 | 0.371 |
| 4.1 - 5.0 years below grade level (%) | 38.9 | 45.6 | -6.7 | 0.057 |
| Sample size | 1,056 | 246 | | |

(continued)

Appendix Table B.4 (continued)

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

NOTES: Baseline data for students in 30 of 34 high schools were collected in spring 2006 (during students' eighth-grade year); baseline data for students in the remaining four schools were collected in fall 2006 (at the start of students' ninth-grade year).

The estimated differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The values in the column labeled "Respondents" are the observed means for students in the study sample with follow-up data on the GRADE assessment (respondents). The "Non-Respondents" values in the next column are the regression-adjusted means for students in the study sample without follow-up data on the GRADE assessment (non-respondents), using the observed distribution of respondents across random assignment blocks (i.e., schools) as the basis for the adjustment.

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

Rounding may cause slight discrepancies in calculating sums and differences.

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

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

characteristics between the ERO and non-ERO groups. The results are presented in Appendix Table B.6. An overall F-test indicates that there is no systematic difference between the two groups either overall or for the RAAL or Xtreme Reading schools.

Comparisons of students in the ERO and non-ERO groups can also be found in Chapter 2. These comparisons are displayed in Table 2.4 for all 34 high schools in the study, in Table 2.5 for the RAAL schools, and in Table 2.6 for the Xtreme Reading schools. Each of these tables indicates a high degree of similarity between students in the ERO and non-ERO groups.

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

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

| Variable | Parameter Estimates (Standard Errors) | | |
|------------------------|--|--------------------------------------|------------------------------|
| | All Schools | Reading Apprenticeship Schools | Xtreme Reading Schools |
| Intercept | 1.360 * | 1.319 * | 1.370 * |
| | (0.322) | (0.442) | (0.470) |
| School 1 | -0.123 | -0.099 | |
| | (0.072) | (0.093) | |
| School 2 | -0.072 | -0.033 | |
| | (0.071) | (0.093) | |
| School 3 | -0.127 | | -0.096 |
| | (0.070) | | (0.072) |
| School 4 | -0.028 | | -0.005 |
| | (0.070) | | (0.072) |
| School 5 | -0.229 * | | -0.193 * |
| | (0.067) | | (0.071) |
| School 6 | -0.137 * | | -0.098 |
| | (0.068) | | (0.071) |
| School 7 | -0.242 * | -0.229 * | |
| | (0.065) | (0.090) | |
| School 8 | -0.120 | -0.107 | |
| | (0.068) | (0.092) | |
| School 9 | -0.046 | | -0.049 |
| | (0.085) | | (0.086) |
| School 10 ^a | -0.040 | -- | |
| | (0.095) | -- | |
| School 11 | -0.127 | -0.111 | |
| | (0.069) | (0.091) | |
| School 12 ^a | -- | | -- |
| | -- | | -- |
| School 13 | -0.138 | | -0.116 |
| | (0.073) | | (0.077) |
| School 14 | -0.214 * | | -0.183 * |
| | (0.076) | | (0.080) |
| School 15 | -0.079 | -0.055 | |
| | (0.070) | (0.093) | |
| School 16 | -0.159 * | -0.132 | |
| | (0.079) | (0.099) | |
| School 17 | -0.106 | -0.074 | |
| | (0.069) | (0.092) | |
| School 18 | -0.137 | | -0.095 |
| | (0.072) | | (0.075) |

(continued)

Appendix Table B.5 (continued)

| Variable | Parameter Estimates (Standard Errors) | | |
|----------------------------|--|--------------------------------------|------------------------------|
| | All Schools | Reading Apprenticeship Schools | Xtreme Reading Schools |
| School 19 | -0.104 (0.069) | | -0.080 (0.072) |
| School 20 | -0.011 (0.074) | 0.005 (0.096) | |
| School 21 | -0.111 (0.067) | -0.067 (0.089) | |
| School 22 | -0.170 * (0.068) | | -0.150 * (0.071) |
| School 23 | -0.098 (0.066) | | -0.076 (0.067) |
| School 24 | -0.098 (0.068) | -0.062 (0.090) | |
| School 25 | -0.075 (0.076) | -0.058 (0.097) | |
| School 26 | -0.040 (0.068) | | -0.004 (0.071) |
| School 27 | -0.112 (0.071) | -0.087 (0.094) | |
| School 28 | -0.096 (0.069) | -0.071 (0.092) | |
| School 29 | -0.143 * (0.070) | | -0.101 (0.074) |
| School 30 | -0.065 (0.070) | | -0.042 (0.074) |
| School 31 | -0.072 (0.064) | -0.060 (0.088) | |
| School 32 | -0.163 * (0.069) | | -0.136 (0.070) |
| School 33 | -0.190 * (0.065) | -0.173 (0.088) | |
| School 34 | -0.157 * (0.070) | | -0.125 (0.072) |
| Research status | | | |
| ERO group | 0.035 * (0.015) | 0.023 (0.021) | 0.046 * (0.022) |
| Non-ERO group ^a | -- -- | -- -- | -- -- |

(continued)

Appendix Table B.5 (continued)

| Variable | Parameter Estimates (Standard Errors) | | |
|--|--|--------------------------------------|------------------------------|
| | All Schools | Reading Apprenticeship Schools | Xtreme Reading Schools |
| Race/ethnicity (%) | | | |
| Hispanic | 0.091 * (0.031) | 0.151 * (0.042) | 0.026 (0.044) |
| Black, non-Hispanic | 0.006 (0.026) | 0.059 (0.037) | -0.048 (0.038) |
| White, non-Hispanic ^a | -- -- | -- -- | -- -- |
| Other | 0.015 (0.036) | 0.111 * (0.050) | -0.088 (0.051) |
| Gender (%) | | | |
| Male | 0.026 (0.015) | 0.016 (0.021) | 0.039 (0.022) |
| Female ^a | -- -- | -- -- | -- -- |
| Average age (years) | -0.059 * (0.020) | -0.069 * (0.028) | -0.050 (0.030) |
| Overage for grade ^b (%) | -0.038 (0.028) | -0.021 (0.039) | -0.059 (0.041) |
| Language other than English spoken at home (%) | -0.021 (0.018) | -0.035 (0.025) | -0.006 (0.026) |
| Language spoken at home missing (%) | -0.267 * (0.083) | -0.357 * (0.108) | -0.131 (0.130) |
| Mother's education level (%) | | | |
| Did not finish high school ^a | -- -- | -- -- | -- -- |
| High school diploma or GED certificate | 0.033 (0.024) | 0.056 (0.033) | 0.003 (0.036) |
| Completed some postsecondary education | 0.035 (0.025) | 0.063 (0.034) | 0.000 (0.037) |
| Don't know | 0.041 (0.026) | 0.090 * (0.036) | -0.014 (0.039) |
| Missing | -0.077 (0.087) | 0.061 (0.118) | -0.258 * (0.130) |
| Father's education level (%) | | | |
| Did not finish high school ^a | -- -- | -- -- | -- -- |
| High school diploma or GED certificate | 0.015 (0.026) | 0.075 * (0.035) | -0.044 (0.037) |
| Completed some postsecondary education | 0.038 (0.027) | 0.068 (0.038) | 0.008 (0.039) |
| Don't know | -0.005 (0.025) | 0.008 (0.035) | -0.021 (0.035) |
| Missing | 0.064 (0.071) | 0.030 (0.094) | 0.091 (0.111) |

(continued)

Appendix Table B.5 (continued)

| Variable | Parameter Estimates (Standard Errors) | | |
|--------------------------------|--|--------------------------------------|------------------------------|
| | All Schools | Reading Apprenticeship Schools | Xtreme Reading Schools |
| GRADE reading comprehension | | | |
| Average standard score | 0.004 * (0.001) | 0.005 * (0.002) | 0.003 (0.002) |
| Sample size | 2,679 | 1,377 | 1,302 |
| Degrees of freedom | 51 | 34 | 34 |
| Mean of the dependent variable | 0.810 | 0.810 | 0.811 |
| R-square | 0.080 | 0.099 | 0.078 |
| F-statistic | 4.492 | 4.331 | 3.161 |
| P-value of F-statistic | 0.000 | 0.000 | 0.000 |

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

NOTES: Baseline data for students in 30 of 34 high schools were collected in spring 2006 (during students' eighth-grade year); baseline data for students in the remaining four schools were collected in fall 2006 (at the start of students' ninth-grade year).

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

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

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

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

| Variable | Parameter Estimates (Standard Errors) | | |
|------------------------|--|--------------------------------------|------------------------------|
| | All Schools | Reading Apprenticeship Schools | Xtreme Reading Schools |
| Intercept | 0.702 (0.480) | 1.216 (0.673) | 0.279 (0.690) |
| School 1 | 0.012 (0.099) | -0.130 (0.130) | |
| School 2 | 0.044 (0.097) | -0.107 (0.129) | |
| School 3 | 0.004 (0.096) | | -0.004 (0.098) |
| School 4 | 0.003 (0.095) | | 0.011 (0.096) |
| School 5 | 0.022 (0.096) | | 0.006 (0.101) |
| School 6 | 0.018 (0.095) | | -0.001 (0.099) |
| School 7 | 0.047 (0.093) | -0.086 (0.128) | |
| School 8 | 0.022 (0.096) | -0.105 (0.130) | |
| School 9 | 0.157 (0.115) | | 0.163 (0.115) |
| School 10 ^a | 0.142 (0.129) | -- -- | |
| School 11 | 0.034 (0.096) | -0.102 (0.129) | |
| School 12 ^a | -- -- | | -- -- |
| School 13 | 0.152 (0.101) | | 0.135 (0.105) |
| School 14 | 0.071 (0.109) | | 0.047 (0.114) |
| School 15 | -0.027 (0.096) | -0.156 (0.129) | |
| School 16 | 0.177 (0.112) | 0.041 (0.141) | |
| School 17 | 0.011 (0.097) | -0.118 (0.129) | |
| School 18 | 0.018 (0.102) | | 0.007 (0.105) |

(continued)

Appendix Table B.6 (continued)

| Variable | Parameter Estimates (Standard Errors) | | |
|----------------------------------|--|--------------------------------------|------------------------------|
| | All Schools | Reading Apprenticeship Schools | Xtreme Reading Schools |
| School 19 | 0.109 (0.097) | | 0.094 (0.099) |
| School 20 | 0.072 (0.101) | -0.060 (0.134) | |
| School 21 | 0.012 (0.092) | -0.126 (0.125) | |
| School 22 | 0.049 (0.094) | | 0.026 (0.097) |
| School 23 | 0.007 (0.090) | | -0.003 (0.091) |
| School 24 | 0.007 (0.092) | -0.117 (0.126) | |
| School 25 | 0.078 (0.105) | -0.066 (0.136) | |
| School 26 | 0.048 (0.093) | | 0.035 (0.096) |
| School 27 | 0.082 (0.097) | -0.051 (0.131) | |
| School 28 | 0.033 (0.093) | -0.095 (0.128) | |
| School 29 | 0.104 (0.097) | | 0.079 (0.102) |
| School 30 | 0.014 (0.095) | | -0.008 (0.100) |
| School 31 | -0.005 (0.088) | -0.138 (0.123) | |
| School 32 | -0.011 (0.096) | | -0.018 (0.098) |
| School 33 | -0.020 (0.092) | -0.169 (0.126) | |
| School 34 | 0.028 (0.098) | | 0.018 (0.101) |
| Race/ethnicity (%) | | | |
| Hispanic | -0.026 (0.044) | -0.029 (0.062) | -0.020 (0.063) |
| Black, non-Hispanic | -0.049 (0.038) | -0.066 (0.055) | -0.025 (0.054) |
| White, non-Hispanic ^a | -- | -- | -- |
| Other | -0.048 (0.052) | -0.015 (0.074) | -0.080 (0.075) |

(continued)

Appendix Table B.6 (continued)

| Variable | Parameter Estimates (Standard Errors) | | |
|--|--|--------------------------------------|------------------------------|
| | All Schools | Reading Apprenticeship Schools | Xtreme Reading Schools |
| Gender (%) | | | |
| Male | -0.036 (0.022) | -0.028 (0.031) | -0.046 (0.031) |
| Female ^a | -- -- | -- -- | -- -- |
| Average age (years) | 0.011 (0.031) | -0.016 (0.043) | 0.043 (0.044) |
| Overage for grade ^b (%) | -0.036 (0.042) | 0.008 (0.058) | -0.086 (0.060) |
| Language other than English spoken at home (%) | -0.024 (0.026) | -0.048 (0.037) | 0.012 (0.038) |
| Home language missing (%) | -0.118 (0.138) | 0.022 (0.200) | -0.259 (0.192) |
| Mother's education level (%) | | | |
| Did not finish high school ^a | -- -- | -- -- | -- -- |
| High school diploma or GED certificate | -0.074 * (0.036) | -0.056 (0.049) | -0.093 (0.053) |
| Completed some postsecondary education | -0.057 (0.036) | -0.034 (0.050) | -0.081 (0.053) |
| Don't know | 0.000 (0.038) | -0.001 (0.053) | 0.006 (0.055) |
| Missing | 0.134 (0.132) | -0.014 (0.181) | 0.288 (0.196) |
| Father's education level (%) | | | |
| Did not finish high school ^a | -- -- | -- -- | -- -- |
| High school diploma or GED certificate | 0.141 * (0.037) | 0.091 (0.053) | 0.191 * (0.053) |
| Completed some postsecondary education | 0.094 * (0.040) | 0.060 (0.057) | 0.130 * (0.055) |
| Don't know | 0.049 (0.036) | -0.005 (0.052) | 0.102 * (0.050) |
| Missing | 0.104 (0.103) | 0.067 (0.140) | 0.150 (0.156) |
| GRADE reading comprehension | | | |
| Average standard score | -0.003 (0.002) | -0.003 (0.003) | -0.004 (0.003) |
| Sample size | 2,171 | 1,115 | 1,056 |
| Degrees of freedom | 50 | 33 | 33 |
| Mean of the dependent variable | 0.582 | 0.578 | 0.586 |
| R-square | 0.023 | 0.018 | 0.038 |
| F-statistic | 0.982 | 0.601 | 1.210 |
| P-value of F-statistic | 0.511 | 0.964 | 0.194 |

(continued)

Appendix Table B.6 (continued)

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

NOTES: Baseline data for students in 30 of 34 high schools were collected in spring 2006 (during students' eighth-grade year); baseline data for students in the remaining four schools were collected in fall 2006 (at the start of students' ninth-grade year).

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

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

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

Appendix C

Statistical Power and Minimum Detectable Effect Size

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

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

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

The ERO impact study was designed to allow an MDES of approximately 0.06 for the full sample of schools in the study and an MDES of approximately 0.10 for the groups of schools using each of the ERO program models. The MDES estimates for the ERO study de-

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

²Lipsey (1990).

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

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

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

⁴Minimum detectable effect sizes were estimated as follows:

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

Where:

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

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

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

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

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

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

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

There are 34 schools in the ERO study sample. Initially, the study aimed to identify approximately 110 students for each of two cohorts of ninth-graders who would be eligible and appropriate for the ERO program. Of these, 60 students would be randomly assigned to enroll in the ERO classes, and the remaining 50 students would constitute the control group. Under these assumptions, the target sample for the second cohort of students in the ERO study was a total of 3,740 students. As discussed in Chapter 2, the actual sample for the second cohort was 2,679 students. This is closer to the sample displayed in the second column of numbers in Appendix Table C.1, which is highlighted to reflect the fact that most of the discussion focuses on the MDES estimates for this sample.

The two remaining columns in Appendix Table C.1 show sample sizes for subgroups comprising 50 percent of the target sample and 25 percent of the sample. The 25 percent subgroup (935 students), for example, is somewhat smaller than the actual number of students in the second cohort with baseline test scores that were between the fourth- and fifth-grade levels (1,092 students.)

The second row of numbers in Appendix Table C.1 shows sample sizes for a subgroup of 17 schools reflecting the groups using each of the two supplemental literacy programs. It shows that the target sample for each ERO program was 1,870 students. In fact, the second cohort includes 1,377 students from the 17 schools using Reading Apprenticeship Academic Literacy (RAAL) and 1,302 students from the 17 schools using Xtreme Reading. These samples are closer to those shown in the second column of numbers in Appendix Table C.1. The third and fourth rows show the sample sizes for smaller subgroups of schools — for example, if the schools within each of the programs were split into two groups (approximately eight schools each) or if there were to be district-level analyses (seven of the 10 participating districts had four schools each).

The bottom panel of Appendix Table C.1 shows sample sizes that would result from follow-up data collection from 80 percent of the students in the original sample. As discussed in Chapter 2, approximately 81 percent of the students in the study sample completed the follow-up test, for a respondent analysis sample of 2,171 students. The resulting samples sizes are closest to those shown in the second column of numbers in Appendix Table C.1.

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

What is the impact of supplemental literacy interventions of the type that were selected on students' reading achievement?

Analyses that address this question rely on the full sample of students across all 34 participating high schools. The second column of numbers in the bottom panel of Appendix Table C.2 indicates that the MDES for this sample would be 0.06 standard deviation if the follow-up data collection effort achieved at least an 80 percent response rate.

What is the impact of each supplemental literacy intervention on students' reading achievement?

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

What is the impact of each supplemental literacy intervention on reading achievement for important subgroups of students or sites?

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

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

The Enhanced Reading Opportunities Study

Appendix Table C.1

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

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

The Enhanced Reading Opportunities Study

Appendix Table C.2

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

| 100 Percent Response Rate Minimum Detectable Effect Size | | | | |
|---|---------------|--------------------------------|--------------------------------|--------------------------------|
| Number of Schools | Target Sample | 75 Percent of Target Sample | 50 Percent of Target Sample | 25 Percent of Target Sample |
| 34 | 0.05 | 0.06 | 0.07 | 0.10 |
| 17 | 0.07 | 0.08 | 0.10 | 0.14 |
| 8 | 0.10 | 0.12 | 0.14 | 0.20 |
| 4 | 0.14 | 0.17 | 0.20 | 0.29 |
| 80 Percent Response Rate Minimum Detectable Effect Size | | | | |
| Number of Schools | Target Sample | 75 Percent of Target Sample | 50 Percent of Target Sample | 25 Percent of Target Sample |
| 34 | 0.06 | 0.06 | 0.08 | 0.11 |
| 17 | 0.08 | 0.09 | 0.11 | 0.16 |
| 8 | 0.11 | 0.13 | 0.16 | 0.23 |
| 4 | 0.16 | 0.19 | 0.23 | 0.32 |

Appendix D

ERO Implementation Fidelity

This appendix describes the development of measures based on the classroom observation data collected during site visits to the high schools in the Enhanced Reading Opportunities (ERO) study. 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 the fall of 2006 and the spring of 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 literacy interventions — Reading Apprenticeship Academic Literacy (RAAL) and Xtreme Reading — and ratings for seven program-specific constructs for each of them. The program-specific constructs reflect the distinctive components of the two literacy 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 2 were conducted by one researcher per school district (a senior staff member with at least a master's degree) and 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).

Site visits were scheduled with the intent of observing classrooms across schools after similar amounts of instructional time had passed. On average, the fall observations occurred 15 weeks after the ERO classes 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 de-

¹Biancarosa and Snow (2004).

velopers 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 — unlike in 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.

- **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 sponta-

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

neous 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, to capitalize on “teachable moments,” or to explain fully a strategy or concept. In these classes, students, while generally attending to the instruction or task at hand, did not appear intellectually engaged, and some may have been inattentive or confused.
- **Category 1.** For each construct, classes that fell into this category were not aligned with the behaviors and activities specified by the developers and described in the protocols. Teachers may have neglected opportunities to teach, may have paid only limited attention to an aspect of the program, and may not have been responsive to students’ confusion or questions. In these classes, students were sporadically engaged in the lesson, and some students may have been acting in a disruptive fashion.

There are five ways in which the study team sought reliable ratings across site visits. First, all observers were trained together to promote a common understanding of the observation process. Second, site visits were conducted by senior study team members, all of whom 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 and reviewers reached a decision on keeping or changing the rating based on review of the observation data. Last, all of the site observers met as a group during the site visits to discuss the rating process and reinforce a common understanding of the relationship between the rating scale and the constructs.

Measuring the Classroom Learning Environment

As discussed in Chapter 3, the measurement of implementation fidelity focused on two key dimensions of implementation: learning environment and comprehension instruction. Ratings for the constructs were combined to calculate composite measures for each of these two key dimensions, for each of the two site visits. This section of the appendix describes how the composite measure of the learning environment dimension was calculated. Because the reliability of these constructs (Cronbach's alpha) is similar across the fall and spring site visits, the reliability is reported for the spring site visit only.

Learning Environment Composite

(2 items, Cronbach's alpha = .86)

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

General Instructional Learning Environment Component

(2 items, Cronbach's alpha = .69)

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

Program-Specific Learning Environment Components

Reading Apprenticeship Academic Literacy (1 item, Cronbach's alpha = NA)

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 (NA).

³In the observation protocols, "motivation and student engagement" is used to describe both a general instructional construct and an Xtreme Reading-specific construct. In this discussion and in Table 3.4, 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."

Xtreme Reading (2 items, Cronbach’s alpha = .88)

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

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

Where:

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

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

Where:

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

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

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 = .79)

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 = .70)

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 = .69)

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: meta-cognitive conversations, silent sustained reading, content/theme integration, writing, and integration of curriculum strands.

Xtreme Reading (2 items, Cronbach's alpha = .63)

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 = .69). The needs-driven instruction construct is the average of two subconstructs: student accommodations and feedback to students (Cronbach's alpha = .51).

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

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

Where:

| | |
|-------------|--|
| CI_{RA} | = comprehension instruction composite measure in a RAAL school |
| GIC_1 | = comprehension (general instructional construct) |
| GIC_2 | = 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) |

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

Where:

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

Categorizing Implementation Fidelity

This section of the appendix discusses briefly how schools were categorized based on the average ratings calculated for each of the 34 participating high schools on the implementation fidelity of their classroom learning environment and for the implementation fidelity of their comprehension instruction. Each 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

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

— 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 especially problematic and where schools’ 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, as discussed in Chapter 3. 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 with weak implementation fidelity; thus, the fidelity groupings used in Year 2 are defined in the same way as in Year 1. That said, only one school was considered poorly aligned on one of the two key dimensions by the end of Year 2 (the spring 2007 site visit).

The ranges of average scores used to define each of the three fidelity groupings are described below. Because the purpose of these groupings was to identify schools whose programs were not representative of the 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.

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 aligned” (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).

Moderately Aligned

The key dimensions were designated as *moderately aligned* in terms of implementation fidelity if the average rating across the general and program-specific constructs used to create the relevant composite was within the range of 1.5 to 1.9. In these cases, the school’s ERO program was observed to have some problems with implementation. In terms of 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.

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.

The top two panels of Appendix Tables D.1 and D.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, moderately aligned, and poorly aligned categories of fidelity during the fall and the spring site visit, respectively. These panels are the same as the top two panels of Table 3.5a and 3.5b in Chapter 3. The bottom panel of these two tables clusters schools based on their level of implementation fidelity across both dimensions. This panel clusters the schools into more categories of combined implementation fidelity than the same panel in Tables 3.5a and 3.5b. Appendix Table D.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. Appendix Tables D.4 and D.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 the fall and spring site visits in Year 2, respectively. Appendix Tables D.6 and D.7 present the distribution across fidelity categories of the 25 schools where the ERO teacher taught two full years of the program, during the spring site visits in Year 1 and Year 2 of the study, respectively.

The Enhanced Reading Opportunities Study

Appendix Table D.1

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 |
|---|-------------|--------------------------------|------------------------|
| <u>Learning environment</u> | | | |
| School average | 2.46 | 2.47 | 2.44 |
| Well-aligned implementation (composite rating is 2.0 or higher) | 30 | 14 | 16 |
| Moderately aligned implementation (composite rating is 1.5-1.9) | 3 | 3 | 0 |
| Poorly aligned implementation (composite rating is less than 1.5) | 1 | 0 | 1 |
| <u>Comprehension instruction</u> | | | |
| School average | 2.10 | 2.10 | 2.10 |
| Well-aligned implementation (composite rating is 2.0 or higher) | 20 | 11 | 9 |
| Moderately aligned implementation (composite rating is 1.5-1.9) | 11 | 5 | 6 |
| Poorly aligned implementation (composite rating is less than 1.5) | 3 | 1 | 2 |
| <u>Combined dimensions</u> | | | |
| School average | 2.28 | 2.28 | 2.27 |
| Well-aligned implementation (composite rating is 2.0 or higher) | 20 | 11 | 9 |
| Well-aligned implementation on learning environment instruction only ^a | 10 | 3 | 7 |
| Well-aligned implementation on comprehension instruction only | 0 | 0 | 0 |
| Moderately aligned implementation on both dimensions | 2 | 2 | 0 |
| Poorly aligned implementation on learning environment only | 0 | 0 | 0 |
| Poorly aligned implementation on comprehension instruction only ^a | 2 | 1 | 1 |
| Poorly aligned implementation on both dimensions | 1 | 0 | 1 |
| Sample size | 34 | 17 | 17 |

(continued)

Appendix Table D.1 (continued)

SOURCES: MDRC and AIR calculations from classroom observation data.

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

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

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

^aOne XR school was designated as being well aligned in terms of learning environment and poorly aligned in terms of comprehension instruction. Thus, this school is counted in two rows in the bottom panel of the table.

The Enhanced Reading Opportunities Study

Appendix Table D.2

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

| Implementation Dimension | All Schools | Reading Apprenticeship Schools | Xtreme Reading Schools |
|--|-------------|--------------------------------|------------------------|
| <u>Learning environment</u> | | | |
| School average | 2.46 | 2.63 | 2.28 |
| Well-aligned implementation (composite rating is 2.0 or higher) | 31 | 17 | 14 |
| Moderately aligned implementation (composite rating is 1.5-1.9) | 2 | 0 | 2 |
| Poorly aligned implementation (composite rating is less than 1.5) | 1 | 0 | 1 |
| <u>Comprehension instruction</u> | | | |
| School average | 2.33 | 2.27 | 2.38 |
| Well-aligned implementation (composite rating is 2.0 or higher) | 28 | 13 | 15 |
| Moderately aligned implementation (composite rating is 1.5-1.9) | 6 | 4 | 2 |
| Poorly aligned implementation (composite rating is less than 1.5) | 0 | 0 | 0 |
| <u>Combined dimensions</u> | | | |
| School average | 2.39 | 2.45 | 2.33 |
| Well-aligned implementation on both dimensions | 26 | 13 | 13 |
| Well-aligned implementation on learning environment instruction only | 5 | 4 | 1 |
| Well-aligned implementation on comprehension instruction only ^a | 2 | 0 | 2 |
| Moderately aligned implementation on both dimensions | 1 | 0 | 1 |
| Poorly aligned implementation on learning environment only ^a | 1 | 0 | 1 |
| Poorly aligned implementation on comprehension instruction only | 0 | 0 | 0 |
| Poorly aligned implementation on both dimensions | 0 | 0 | 0 |
| Sample size | 34 | 17 | 17 |

(continued)

Appendix Table D.2 (continued)

SOURCES: MDRC and AIR calculations from classroom observation data.

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

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

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

^a One XR school was designated as being well aligned in terms of comprehension instruction and poorly aligned in terms of learning environment. Thus, this school is counted in two rows in the bottom panel of the table.

The Enhanced Reading Opportunities Study

Appendix Table D.3

Number of ERO Classrooms with Well-, Moderately, or Poorly Aligned Implementation to Program Models on Each Implementation Dimension, by ERO Program — Year 2 Spring Site and Fall Visits

| Implementation Dimension | All Schools | Reading Apprenticeship Schools | Xtreme Reading Schools |
|--|-------------|--------------------------------|------------------------|
| <u>Learning environment</u> | | | |
| School average | 2.46 | 2.55 | 2.36 |
| Well-aligned implementation (composite rating is 2.0 or higher) | 31 | 15 | 16 |
| Moderately aligned implementation (composite rating is 1.5-1.9) | 2 | 2 | 0 |
| Poorly aligned implementation (composite rating is less than 1.5) | 1 | 0 | 1 |
| <u>Comprehension instruction</u> | | | |
| School average | 2.21 | 2.18 | 2.24 |
| Well-aligned implementation (composite rating is 2.0 or higher) | 23 | 11 | 12 |
| Moderately aligned implementation (composite rating is 1.5-1.9) | 10 | 5 | 5 |
| Poorly aligned implementation (composite rating is less than 1.5) | 1 | 1 | 0 |
| <u>Combined dimensions</u> | | | |
| School average | 2.33 | 2.37 | 2.30 |
| Well-aligned implementation on both dimensions | 23 | 11 | 12 |
| Well-aligned implementation on learning environment instruction only | 8 | 4 | 4 |
| Well-aligned implementation on comprehension instruction only ^a | 0 | 0 | 0 |
| Moderately aligned implementation on both dimensions | 1 | 1 | 0 |
| Poorly aligned implementation on learning environment only | 1 | 0 | 1 |
| Poorly aligned implementation on comprehension instruction only ^a | 1 | 1 | 0 |
| Poorly aligned implementation on both dimensions | 0 | 0 | 0 |
| Sample size | 34 | 17 | 17 |

(continued)

Appendix Table D.3 (continued)

SOURCES: MDRC and AIR calculations from classroom observation data.

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

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

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

The Enhanced Reading Opportunities Study
Appendix Table D.4
Average Implementation Composite Scores,
by ERO Program — Year 2 Fall

| Characteristic | All Schools | Reading Apprenticeship Schools | Xtreme Reading Schools |
|---|----------------|--------------------------------------|------------------------------|
| <u>Learning environment</u> | | | |
| Schools with replacement teachers | 2.5 | 2.6 | 2.5 |
| Schools with teachers who taught two full years | 2.4 | 2.4 | 2.4 |
| <u>Comprehension instruction</u> | | | |
| Schools with replacement teachers | 2.2 | 2.2 | 2.1 |
| Schools with teachers who taught two full years | 2.1 | 2.1 | 2.1 |
| <u>Combined dimensions</u> | | | |
| 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 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.

The Enhanced Reading Opportunities Study
Appendix Table D.5
Average Implementation Composite Scores,
by ERO Program — Year 2 Spring

| Characteristic | All Schools | Reading Apprenticeship Schools | Xtreme Reading Schools |
|---|----------------|--------------------------------------|------------------------------|
| <u>Learning environment</u> | | | |
| Schools with replacement teachers | 2.4 | 2.6 | 2.3 |
| Schools with teachers who taught two full years | 2.5 | 2.6 | 2.3 |
| <u>Comprehension instruction</u> | | | |
| Schools with replacement teachers | 2.3 | 2.3 | 2.2 |
| Schools with teachers who taught two full years | 2.3 | 2.3 | 2.4 |
| <u>Combined dimensions</u> | | | |
| 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 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.

The Enhanced Reading Opportunities Study

Appendix Table D.6

Number of ERO Classrooms Taught by Teachers Who Taught Two Full Years with Well-, Moderately, or Poorly Aligned Implementation to Program Models on Each Implementation Dimension, by ERO Program — Year 1 Spring

| Implementation Dimension | All Schools | Reading Apprenticeship Schools | Xtreme Reading Schools |
|--|-------------|--------------------------------|------------------------|
| <u>Learning environment</u> | | | |
| Well-aligned implementation (composite rating is 2.0 or higher) | 22 | 11 | 11 |
| Moderately aligned implementation (composite rating is 1.5-1.9) | 3 | 2 | 1 |
| Poorly aligned implementation (composite rating is less than 1.5) | 0 | 0 | 0 |
| <u>Comprehension instruction</u> | | | |
| Well-aligned implementation (composite rating is 2.0 or higher) | 13 | 5 | 8 |
| Moderately aligned implementation (composite rating is 1.5-1.9) | 8 | 4 | 4 |
| Poorly aligned implementation (composite rating is less than 1.5) | 4 | 4 | 0 |
| <u>Combined dimensions</u> | | | |
| Well-aligned implementation on both dimensions | 13 | 5 | 8 |
| Moderately aligned implementation on at least one dimension and moderately or well-aligned implementation on the other dimension | 8 | 4 | 4 |
| Poorly aligned implementation on at least one dimension | 4 | 4 | 0 |
| Sample size | 25 | 13 | 12 |

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.

The Enhanced Reading Opportunities Study

Appendix Table D.7

Number of ERO Classrooms Taught by Teachers Who Taught Two Full Years 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 |
|--|-------------|--------------------------------|------------------------|
| <u>Learning environment</u> | | | |
| Well-aligned implementation (composite rating is 2.0 or higher) | 23 | 13 | 10 |
| Moderately aligned implementation (composite rating is 1.5-1.9) | 1 | 0 | 1 |
| Poorly aligned implementation (composite rating is less than 1.5) | 1 | 0 | 1 |
| <u>Comprehension instruction</u> | | | |
| Well-aligned implementation (composite rating is 2.0 or higher) | 22 | 10 | 12 |
| Moderately aligned implementation (composite rating is 1.5-1.9) | 3 | 3 | 0 |
| Poorly aligned implementation (composite rating is less than 1.5) | 0 | 0 | 0 |
| <u>Combined dimensions</u> | | | |
| Well-aligned implementation on both dimensions | 20 | 10 | 10 |
| Moderately aligned implementation on at least one dimension and moderately or well-aligned implementation on the other dimension | 4 | 3 | 1 |
| Poorly aligned implementation on at least one dimension | 1 | 0 | 1 |
| Sample size | 25 | 13 | 12 |

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.

Classroom Observation Scales

Enhanced Reading Opportunities Study

American Institutes for Research

November 2006

Table of Contents

| | |
|--------------------------------------|-----|
| General Instruction Scales..... | 186 |
| RAAL Fidelity Measure..... | 191 |
| Xtreme Reading Fidelity Measure..... | 199 |

Enhanced Reading Opportunities Program
General Instruction Scales

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

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

| Area of interest | Vocabulary |
|--|------------|
| Description | |
| <ol style="list-style-type: none"><li data-bbox="233 373 1308 436">0. Not applicable. There was no opportunity for vocabulary instruction to occur during the observed class period(s).<li data-bbox="233 457 1414 617">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.<li data-bbox="233 638 1406 798">2. Students are engaged in some vocabulary activities, but these activities are not fully developed. For example, the teacher may employ definitional and contextual information for presenting words but gives little attention to linking words to prior experiences OR to teaching strategies to help students figure out the meaning of words on their own (e.g. identifying root word, using context clues, etc).<li data-bbox="233 819 1373 1016">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 | |
| <ol style="list-style-type: none"> <li data-bbox="233 329 1417 527">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. <li data-bbox="233 548 1417 842">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. <li data-bbox="233 863 1417 1228">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.) | |
| <ol style="list-style-type: none"> <li data-bbox="233 401 1406 632">1. Little metacognitive work is apparent, and overall, metacognitive skills are not being <i>developed</i> through instruction or conscious practice. In some cases, students may be taught strategies to monitor their own reading, recognize faulty comprehension, and apply “fix-up” strategies; but these strategies are not explored. For example, the teacher either does not address metacognitive strategies (e.g., self-monitoring of reading may not be taught at all) or does so in a very limited or superficial, contrived manner (e.g., teacher and students are most often “going through the motions”). <li data-bbox="233 653 1419 848">2. Instruction incorporates some development of metacognitive strategies and opportunities for student practice of metacognition, either through spoken or written expression, but these may not be fully developed. For example, instruction could include the use of “think alouds” to model strategies, self-correct, and make connections to prior knowledge. While some of the metacognitive activities flow naturally, others may appear to be forced (teacher or students appear to be “going through the motions”). <li data-bbox="233 869 1414 1064">3. Use of metacognitive strategies is pervasive and integrated throughout instruction. Instruction includes teacher modeling of strategies and multiple opportunities for student practice of thinking aloud through spoken or written expression with multiple forms of text. Throughout the majority of metacognitive activities, the teacher monitors and guides students in their thought processes. In addition, the majority of the metacognitive activities are conducted in a natural and thoughtful manner. | |

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

| Area of interest | Motivation and Student Engagement |
|---|-----------------------------------|
| Description | |
| <ol style="list-style-type: none"> <li data-bbox="233 1148 1411 1274">1. Disruptive or passive disengagement; most students are frequently off-task, as evidenced by either gross inattention or serious disruptions. For substantial portions of time, many students are either off-task or nominally on-task but not trying very hard. Students could appear to be lethargic and disinterested in class activities or they might be actively misbehaving. <li data-bbox="233 1299 1382 1390">2. Sporadic or episodic engagement; most students, some of the time, are engaged in class activities. Engagement may be uneven, mildly enthusiastic or dependent on frequent prodding from the teacher. <li data-bbox="233 1415 1411 1476">3. Engagement is widespread; most students are on-task most of the time pursuing the substance of the lesson. The majority of students seem to be taking the work seriously and trying hard. | |

Enhanced Reading Opportunities Program

Reading Apprenticeship Academic Literacy Fidelity Scales

Core Principle # 1

Social Reading Community

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

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

Fidelity Scale

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

Such an environment could be characterized by little or no student sharing related to the evaluation or generation of meaning from text. Many students may appear to be reluctant to participate in discussions related to text most of the time. The teacher may have to work extremely hard to get students to interact about text meaning, or prompting by the teacher to encourage student conversations about literature is ineffective.

Instruction in this category could also be characterized by students ridiculing their peers when they acknowledge confusion about text. The teacher may ignore student attempts to express confusion or may not model respect for the varied perspectives and ideas of all members of the classroom community.

2. In general, the classroom environment appears to be a safe place to interact and share ideas about text. The teacher occasionally models appropriate ways for sharing ideas about text.

A moderately developed social reading community could be characterized by discussions about text that are primarily teacher-directed during the majority of the instructional period. Classroom routines for peer interaction may not be fully developed. Some students may appear to be hesitant to volunteer their own ideas or confusion about text. As another example, the teacher may actively listen to student responses and attempt to elicit a variety of responses from all members of the reading community, but he or she has trouble engaging the majority of students in discussion of literature or of text meaning.

3. A safe and nurturing environment is established for students to share ideas about text. When necessary, the teacher models a process for sharing ideas about text.

This social reading community could be characterized by frequent student participation. The majority of students contribute to or guide whole-class or group conversations and activities related to literature and other forms of text. They may also volunteer confusion and difficulties with texts. A positive social reading community could also be evident during teacher-facilitated conversations that encourage active participation from all members of the classroom community.

| | |
|---------------------------|-----------------------------------|
| Core Principle # 2 | Metacognitive Conversation |
|---------------------------|-----------------------------------|

Metacognitive Conversation is a regularly occurring routine which is evident in RAAL classroom work and interactions:

- Students are taught to use classroom inquiry to generate a repertoire of specific comprehension and problem-solving strategies.
- Through ongoing conversations rooted in text, students learn to ask critical questions about content, purpose, and perspective.
- Students are encouraged to draw on strategic skills they use in out-of-school settings to assist them in solving comprehension problems.
- Students recognize that confusion can be a starting place for collaborative problem-solving aimed at deriving meaning from difficult text.
- Students have many opportunities to practice sharing and exploring their thinking about texts *with peers*; these peer-guided metacognitive conversations become more text-based and sophisticated over the course of the academic year.
- Students monitor their own mental processes for reading and adjust as needed.*
- During discussions, teacher probes for deeper student responses to enrich student learning and thinking processes.
- Teacher models metacognitive process (e.g. Thinking Aloud, Talking to the Text) *and* follows through on such practices with continued modeling and appropriate scaffolding to ensure that streams of thought are fully developed.

| |
|-----------------------|
| Fidelity Scale |
|-----------------------|

1. Students are not explicitly taught a variety of comprehension and problem-solving skills. Students are primarily engaged in instruction that is aimed at uniform understandings and single correct responses.

For example, there is little evidence that reading comprehension difficulties are seen as valuable starting points for collaborative problem-solving. Students have few opportunities to practice discussing their thought processes about reading and to ask critical questions about text content. Students do not volunteer to discuss confusion about text. Students are never or rarely asked to make connections to strategic skills they use in out-of-school settings to assist them in solving comprehension problems.

As another example, the teacher does not model metacognitive strategies, or does not provide scaffolds for students to practice and apply such strategies. Instruction that falls into this category could be characterized by teacher attempts to model the use of metacognitive strategies that are largely unsuccessful or ineffective.

2. Students are taught comprehension and problem-solving skills, and at least one major classroom activity provides students with an opportunity to discuss their cognitive processes.

For example, some but not all students may share reading difficulties and confusions and collaborate in problem solving. Instruction could include opportunities for students to share problem solving and strategic skills from their lives outside of school.

Instruction could also include teacher or student engagement in discussion or assessment of the effects of particular reading processes. While the teacher occasionally models metacognitive strategies or probes for deeper student responses in relation to text, only minimal attempts are made to follow through with additional modeling or appropriate scaffolds to ensure that thought streams are fully developed and transparent.

* While we are including this bullet in the general description of the principles, we will not include in the fidelity scales as this is a “high inference” item and is not easily observable.

3. Students are taught a variety of comprehension and problem-solving skills, and they actively contribute to or guide metacognitive conversations. Such conversations are predominantly text-based.

For example, many students routinely make connections to strategic skills they use in out-of-school settings to assist them in solving comprehension problems. Students may also share their confusion with text as a basis for comprehending challenging text.

As another example, the teacher frequently and authentically models metacognitive strategies (such as using confusion as a point to generate meaning) or probes for deeper student responses in relation to text. Initial modeling is followed by additional modeling and/or appropriate scaffolds aimed at ensuring that thought streams are fully developed and transparent.

Core Principle # 3**Silent Sustained Reading**

Silent Sustained Reading is a well-established routine in which personal inquiry and peer social interaction is used to build motivation and extend students' interest to new books and genres.

- Students are encouraged to explore their own preferences and reactions to books.
- Students routinely discuss SSR books with classmates in both informal and occasionally formal activities (i.e. “book talks”).
- Students set goals for their reading development and assess their own performance in meeting those goals (in terms of amount and range of books read, persistence, and fluency).
- Students practice metacognitive routines, language study, and cognitive strategies as they read SSR books.
- Teachers routinely provide support and show interest in students’ SSR in both informal and formal activities, e.g., individual conferencing, written feedback in reading logs, sharing their own SSR books and reading processes.

Fidelity Scale

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) |
|---|

- | |
|--|
| <ol style="list-style-type: none"> 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 E

Technical Notes for Impact Findings

This appendix provides three sets of additional technical notes that accompany the impact findings presented in Chapter 5. The first section describes the statistical model used to estimate the impact of the two supplemental literacy programs used in the Enhanced Reading Opportunities (ERO) study — Reading Apprenticeship Academic Literacy (RAAL) and Xtreme Reading — on reading achievement and reading behaviors. The second section presents tables that show the sensitivity of the core impact findings to including student-level baseline characteristics in the statistical model, for the full sample of 34 schools and for the groups of schools using each of the two literacy programs. These tables also present the standard errors (“S.E.” in the tables) and 95 percent confidence intervals for the adjusted and unadjusted impacts. The third section addresses the issues related to multiple hypothesis tests of impacts on multiple reading behavior measures. Specifically, it presents the findings from the qualifying tests that were performed to assess the robustness of the statistical significance of the impacts on the three reading behavior measures examined in Chapter 5.

Statistical Model for Estimating Impacts

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

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

Where:

Y_i = reading achievement or reading behaviors outcome for student i

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

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

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

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

ε_i = student-level random error term

In this model, β_0 represents the estimated impact of the ERO programs on the outcome of interest (Y_i). β_0 is a fixed-effect impact estimate that addresses the question: What is the impact of the ERO programs for the average student in the follow-up respondent sample? This approach is taken because this study most closely reflects an efficacy study of the effects of a

new supplemental literacy intervention under relatively controlled conditions. Also, the sites and students were not selected to be a random sample of a larger population of sites. Instead, sites were selected purposively through the OVAE special SLC grant competition discussed in Chapter 2, using specific criteria that differentiated these schools and districts from others that were not awarded a grant. Although, on average, the participating schools share characteristics of other low-performing urban high schools across the country, the impact estimates are not statistically generalizable to a larger population of districts, high schools, or students.

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

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

Adjusted and Unadjusted Impact Estimates

As explained above, the impacts presented in Chapter 5 of this report are estimated using regression adjustments for random differences between the ERO and non-ERO groups in their pretest scores and whether a student was overage for the ninth grade. The first two tables in this appendix provide both regression-adjusted (in the "Estimated Impact" column) and unadjusted impacts (in the "Difference" column). These tables also include other information that may be useful to those who may wish to include these impacts in meta-analyses. Note that random assignment of students to the ERO and non-ERO groups occurred within each high school (that is, random assignment was "blocked" by school). Because of differences across schools (blocks) in the number of students eligible and appropriate for the ERO programs, the ratio of ERO group members to non-ERO group members in each site varies from 1.14 to 2.0. Thus, all the impact estimates presented in this report include controls for each block to account for random differences between the ERO and non-ERO groups that may be associated with differences in the random assignment ratios. The assessment of sensitivity to other regression adjustments presented in the appendix reflects potential differences in impact estimates that also controls for the blocking of random assignment by school.

Appendix Table E.1 is the counterpart to Tables 5.1 and 5.2 and shows adjusted and unadjusted impacts on reading achievement for all 34 schools in the study and for the groups of

The Enhanced Reading Opportunities Study
Appendix Table E.1
Impacts on Reading Achievement,
Cohort 2 Follow-Up Respondent Sample

| Outcome | Mean Differences Adjusting for Blocking Only | | | | Regression-Based Impact Estimates | | | | |
|---|--|----------------------|--------------------------------|----------------------------|-----------------------------------|-------------------------|-----------------------------|-----------------|-------------------------|
| | ERO Group (S.D.) | Non-ERO Group (S.D.) | 95% Confidence Interval (S.E.) | P-Value for the Difference | Estimated Impact (S.E.) | 95% Confidence Interval | P-Value of Estimated Impact | Impact (S.E.) | 95% Confidence Interval |
| <u>All schools</u> | | | | | | | | | |
| Reading comprehension Average standard score | 90.10 (10.15) | 89.49 (10.03) | 0.61 -0.25 (0.44) | 1.46 0.166 | 0.83 (0.41) | 0.03 1.63 | 0.042 | 0.08 (0.04) | 0.00 0.16 |
| Reading vocabulary Average standard score | 93.45 (10.34) | 93.68 (9.83) | -0.22 (0.43) | -1.07 0.63 0.607 | -0.01 (0.41) | -0.80 0.79 | 0.986 | 0.00 (0.04) | -0.08 0.08 |
| Sample size | 1,264 | 907 | | | | | | | |
| <u>Reading Apprenticeship schools</u> | | | | | | | | | |
| Reading comprehension Average standard score | 90.24 (10.02) | 89.05 (10.11) | 1.19 (0.61) | 0.00 2.38 0.049 | 1.38 (0.57) | 0.27 2.50 | 0.015 | 0.14 (0.06) | 0.03 0.25 |
| Reading vocabulary Average standard score | 93.39 (10.04) | 94.02 (9.31) | -0.62 (0.58) | -1.77 0.52 0.285 | -0.43 (0.55) | -1.51 0.64 | 0.428 | -0.04 (0.06) | -0.15 0.07 |
| Sample size | 645 | 470 | | | | | | | |
| <u>Xtreme Reading schools</u> | | | | | | | | | |
| Reading comprehension Average standard score | 89.96 (10.30) | 89.97 (9.95) | -0.02 (0.63) | -1.26 1.22 0.980 | 0.25 (0.59) | -0.91 1.41 | 0.672 | 0.02 (0.06) | -0.09 0.14 |
| Reading vocabulary Average standard score | 93.52 (10.66) | 93.32 (10.35) | 0.20 (0.64) | -1.06 1.46 0.752 | 0.44 (0.60) | -0.75 1.62 | 0.468 | 0.04 (0.06) | -0.08 0.17 |
| Sample size | 619 | 437 | | | | | | | |

(continued)

Appendix Table E.1 (continued)

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

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

The estimated impacts in the column labeled "Difference" are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. 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 (i.e., schools) as the basis for the adjustment. The estimated impacts in the column labeled "Estimated Impact" are regression-adjusted for blocking of random assignment by school, as well as for random differences between the ERO and non-ERO groups in their baseline reading comprehension test scores and age at random assignment.

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

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

A two-tailed t-test was applied to the impact estimate. The statistical significance is indicated (*) when the p-value is less than or equal to 5 percent. Rounding may cause slight discrepancies in calculating sums and differences.

schools using each of the two ERO programs. Appendix Table E.2 is the counterpart to Tables 5.3 and 5.4 and shows adjusted and unadjusted impacts on reading behavior measures.¹

Addressing Risks Associated with Multiple Hypothesis Tests

In Chapter 5, statistical significance is indicated in the tables by an asterisk (*) when the p-value of the impact estimate is less than or equal to 0.05 (5 percent). As discussed in Chapter 2, however, when making judgments about statistical significance, it is important to recognize potential problems associated with conducting multiple hypothesis tests. Specifically, it is important to minimize the risk that conclusions from the study could be based on false positive results (also known as Type I errors) while simultaneously limiting the risk that important results may be neglected due to false negative results (also known as Type II errors). In other words, the analysis should avoid concluding that an impact estimate is statistically significant when, in fact, there is no true impact. Likewise the analysis should not be so conservative with respect to producing false positives that it unduly increases the likelihood of missing true impacts when they exist (that is, of producing false negatives).

As the number of hypothesis tests increases, the probability of finding a statistically significant impact estimate when there is no true impact may also increase. One could dramatically reduce this risk by making the standard for statistical significance much more stringent, for example, by setting the p-value to less than or equal to 0.001. Making the standard too stringent, however, will increase the likelihood that one would judge an impact estimate to be not statistically significant when, in fact, it represents a true impact. The approach adopted for this project provides a framework that aspires for an *acceptable balance* between the risks of making Type I and Type II errors.

The impact analysis conducted for this report includes two sets of safeguards aimed at attenuating the risk of drawing inappropriate conclusions about program effectiveness on the basis of multiple hypothesis tests. The first safeguard is to identify a *parsimonious* list of outcome measures and subgroups and then to prioritize among these to specify the primary and secondary hypothesis tests that would be used to make judgments about the overall effectiveness of the ERO programs. The shorter this list, the fewer the number of hypothesis tests and, thus, the less exposed the analysis will be to “spurious statistical significance” as a result of having tested multiple hypotheses.

¹Results from the regression-adjusted impact analyses are presented in the columns under “Regression-Based Impact Estimates,” and results from the unadjusted impact analyses are presented in the columns under “Mean Differences Adjusted for Blocking Only.”

The Enhanced Reading Opportunities Study
Appendix Table E.2
Impacts on Reading Behaviors,
Cohort 2 Follow-Up Respondent Sample

| Outcome | Mean Differences Adjusting for Blocking Only | | | | Regression-Based Impact Estimates | | | | | |
|--|--|----------------------|--------------------------------|----------------------------|-----------------------------------|-------------------------|-----------------------------|---------------|-------------------------|--------------------|
| | ERO Group (S.D.) | Non-ERO Group (S.D.) | 95% Confidence Interval (S.E.) | P-Value for the Difference | Estimated Impact (S.E.) | 95% Confidence Interval | P-Value of Estimated Impact | Impact (S.E.) | 95% Confidence Interval | |
| All schools | | | | | | | | | | |
| Amount of school-related reading (prior month occurrences) | 46.29 (36.93) | 46.03 (38.32) | 0.25 (-1.60) | 3.38 | 0.875 | 0.40 (-1.59) | -2.72 | 3.53 | 0.800 (0.04) | 0.01 (-0.07) 0.09 |
| Amount of non-school-related reading (prior month occurrences) | 29.88 (32.36) | 29.38 (32.98) | 0.50 (-1.41) | 3.26 | 0.722 | 0.49 (-1.41) | -2.28 | 3.25 | 0.729 (0.04) | 0.01 (-0.07) 0.10 |
| Use of reflective reading strategies (4-point scale) | 2.76 (0.59) | 2.71 (0.59) | 0.05 (0.03) | 0.10 | 0.047 | 0.05 (0.03) | 0.00 | 0.10 | 0.033 (0.04) | 0.09 (0.01) 0.17 |
| Sample size | 1,260 | 901 | | | | | | | | |
| Reading Apprenticeship schools | | | | | | | | | | |
| Amount of school-related reading (prior month occurrences) | 50.46 (38.58) | 48.85 (39.97) | 1.61 (-2.34) | 6.20 | 0.492 | 1.64 (-2.34) | -2.95 | 6.23 | 0.483 (0.06) | 0.04 (-0.08) 0.16 |
| Amount of non-school-related reading (prior month occurrences) | 31.30 (31.73) | 31.52 (35.10) | -0.22 (-2.00) | 3.71 | 0.913 | -0.34 (-2.00) | -4.27 | 3.59 | 0.867 (0.06) | -0.01 (-0.13) 0.11 |
| Use of reflective reading strategies (4-point scale) | 2.78 (0.59) | 2.74 (0.57) | 0.04 (0.03) | 0.10 | 0.297 | 0.04 (0.03) | -0.03 | 0.11 | 0.257 (0.06) | 0.07 (-0.05) 0.18 |
| Sample size | 642 | 466 | | | | | | | | |
| Xtreme Reading schools | | | | | | | | | | |
| Amount of school-related reading (prior month occurrences) | 41.97 (34.65) | 43.16 (36.29) | -1.20 (-2.16) | 3.04 | 0.579 | -0.84 (-2.15) | -5.05 | 3.38 | 0.697 (0.06) | -0.02 (-0.13) 0.09 |
| Amount of non-school-related reading (prior month occurrences) | 28.41 (32.97) | 27.14 (30.41) | 1.27 (-1.97) | 5.14 | 0.521 | 1.46 (-1.97) | -2.41 | 5.34 | 0.458 (0.06) | 0.04 (-0.07) 0.16 |
| Use of reflective reading strategies (4-point scale) | 2.74 (0.59) | 2.68 (0.61) | 0.06 (0.04) | 0.14 | 0.079 | 0.07 (0.04) | 0.00 | 0.14 | 0.059 (0.06) | 0.12 (0.00) 0.24 |
| Sample size | 618 | 435 | | | | | | | | |

(continued)

Appendix Table E.2 (continued)

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

NOTES: The student follow-up survey was administered in spring 2007 at the end of students' ninth-grade year.

The estimated impacts in the column labeled "Difference" are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. 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 (i.e., schools) as the basis for the adjustment. The estimated impacts in the column labeled "Estimated Impact" are regression-adjusted for blocking of random assignment by school, as well as for random differences between the ERO and non-ERO groups in their baseline reading comprehension test scores and age at random assignment.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 38.322; non-school-related reading standard deviation = 32.976; use of reading strategies standard deviation = 0.592).

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 3 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

The second safeguard uses composite statistical tests to “qualify” or call into question multiple hypothesis tests that are statistically significant individually but that may be due to chance. These composite tests are referred to as “qualifying tests.”

Specifying Primary and Secondary Hypothesis Tests

The primary evidence of overall ERO program effectiveness for this report will be reflected by estimates of program impacts on reading comprehension test scores (expressed in standard score values) for the full study sample and for each of the two ERO programs being evaluated. Anchoring the study’s early conclusions in a limited set of outcomes minimizes the risk of relying on a large number of impact estimates, some of which may be statistically significant only by chance. As noted above, student reading comprehension skills constitute the primary target of the ERO interventions and the primary outcome of interest for the first year of the study. Also, the study was designed to provide minimum detectable effect sizes for each ERO subgroup that may be considered policy relevant. Thus, the primary confirmatory hypotheses for the report focus on the overall and program-specific impacts on reading comprehension test scores.

Vocabulary knowledge and student reading behaviors, while targets of the interventions and important to students’ literacy development, are considered secondary indicators of program effectiveness. Similarly, subgroups of students (for example, those with higher or lower baseline test scores) provide useful information about the relative impact of supplemental literacy programs, but they too are considered secondary indicators of effectiveness in this report.

Composite Qualifying Statistical Tests

A second set of safeguards against risks associated with multiple hypothesis tests involves the use of composite qualifying statistical tests that provide further context for interpreting the robustness of individual impact estimates and their statistical significance.² These statistical tests are applied in cases where impacts are estimated for more than one outcome in a given measurement domain (for example, the three survey measures that attempt to capture students’ reading behaviors) or for subgroups of the full study sample. In general, these qualifying statistical tests estimate impacts on composite indices that encompass all the measures in a given domain or estimate the overall variation in impacts across subgroups. If the results of these tests are not statistically significant, this indicates that the statistical significance of the associated

²Measurement of overall effects has its roots in the literature on meta-analysis (see O’Brien, 1984; Logan and Tamhane, 2003; and Hedges and Olkin, 1985). For a discussion of qualifying statistical tests to account for the risk of Type I error, see Duflo, Glennerster, and Kremer (2007). Other applications of these approaches are discussed in Kling and Liebman (2004) and Kling, Liebman, and Katz (2007).

individual impact estimates *may have* occurred by chance. In these cases, the discussion of the impacts should include cautions or qualifiers about the robustness of the individual findings.³

To test the robustness of the statistical significance of impact estimates for multiple outcomes within a measurement domain (in this case, the three reading behavior measures), the study uses a single composite index consisting of the average of the standardized values for each outcome.⁴ Then the estimated impact on this composite measure is calculated for the full study sample. If this qualifying test shows that the composite impact estimate is not statistically significant (its p-value is greater than 0.05), then one concludes that statistically significant impacts for the component outcomes could be due to chance and should be interpreted cautiously.

Specifically, the analysis took the following steps in creating a composite index and assessing impacts on reading behaviors.⁵ First, z-scores were created for each reading behavior outcome by subtracting the non-ERO group mean and dividing by the non-ERO group standard deviation. Thus, each component of the index has a mean of zero and a standard deviation of one for the non-ERO group. The z-scores from each component were averaged to obtain the index which was then included in the standard impact estimation model. If the estimated impact for the composite index is not statistically significant, then the statistical significance of impact estimates for the component measures may have occurred by chance and the finding should be interpreted cautiously. In other words, the report qualifies or calls into question a statistically significant individual impact estimate by suggesting that it may have occurred by chance.

To test the robustness of the statistical significance of impact estimates for subgroups of students, a composite F-test is used to assess whether the variation in impacts across all student

³Alternative strategies that involve (1) adjusting significance levels (through Bonferroni methods) or (2) adjusting significance thresholds (through Benjamini and Hochberg methods) are overly conservative with respect to making Type I errors and can thereby greatly increase the likelihood of making Type II errors. There are two reasons for this. First, these methods treat all hypotheses as though they were independent of each other. Hence, each hypothesis is treated as representing an independent opportunity to make a Type I error. However, many impact estimates in an evaluation study are correlated with each other and thus do not represent independent opportunities to make Type I errors. In the extreme, for example, if all measures were perfectly correlated, there is only one opportunity to make a Type I error even though there are many outcome measures and, thus, many statistical hypothesis tests. The above methods assume, however, that the number of opportunities to make a Type I error equals the number of hypothesis tests conducted. To the degree that hypothesis tests are correlated with each other, these methods overcompensate (often by a lot) for the risks of Type I error in multiple hypothesis tests. A second source of conservatism with respect to Type I error is the fact that the above methods assume that all null hypotheses may be true. As a result, they consider the potential number of false positives to equal the total number of hypothesis tests conducted. However, the actual number of potential false positives equals the total number of true null hypotheses, not the total number of hypotheses tested. This is because only true null hypotheses can produce false positives. Hence, the methods overcompensate for the number of hypotheses tested.

⁴See Duflo, Glennerster, and Kremer (2007).

⁵The discussion and method presented here draw from Kling, Liebman, and Katz (2007).

subgroups is statistically significant. For example, the analysis examines impacts for three sets of student subgroups: those defined by baseline reading test scores (comprising three subgroups); those defined by whether a student was overage for the start of ninth grade (comprising two subgroups); and those defined by whether a student's family spoke a language other than English at home (comprising two subgroups). The composite qualifying test for these analyses assesses whether variation in estimated impacts across these seven subgroups accounts for a statistically significant level of unexplained variance in the test score or other outcome being examined. In other words, the test assesses whether the change in the F-statistic from the core impact regression to the impact regression with the subgroup interaction terms is statistically significant (its p-value is less than or equal to 0.05). If the change in unexplained variance due to the subgroup impact interactions is not statistically significant, then the statistical significance of impact estimates for the component subgroups may have occurred by chance and the findings should be interpreted cautiously.

Finally, the analysis includes qualifying statistical tests to assess the statistical significance of the difference in impacts between the subgroups of students or schools. If these qualifying tests show that the difference in impacts across subgroups is not statistically significant (p-value is greater than 0.05), then one concludes that statistically significant impacts for individual subgroups could be due to chance and should be interpreted cautiously.⁶ For example, suppose the findings indicate that impacts on reading comprehension for one group of participating high schools are positive and statistically significant while the result for a second group of schools is also positive but is not statistically significant. If the difference in impacts between the two groups of schools is not statistically significant, one should be especially cautious about concluding that the ERO programs were more effective for some schools than for others.

Appendix Table E.3 displays the results of the composite qualifying statistical tests for the three reading behavior measures discussed in Chapter 5. As discussed above, the composite index was created by averaging the standardized values of the three reading behaviors outcomes: amount of school-related reading, amount of non-school-related reading, and use of reflective reading strategies. Appendix Table E.3 shows results for the full sample of all schools, for each of the two ERO programs separately, and for the various subgroups that are discussed in Chapter 5. None of the estimated impacts on the composite index is statistically significant at the 5 percent level. Thus, readers should exercise caution in interpreting statistically significant impacts for the individual components of the composite index, since these may be due to chance. Appendix Table E.3 also includes the results of the composite qualifying statistical test of the robustness of statistical significance of the *difference in impacts* across subgroups of students or schools. It shows that the difference in impacts is also not statistically significant for the full sample or any of the subgroups. Thus, the difference in impacts should be interpreted cautiously.

⁶Note that one conducts qualifying statistical tests using the composite index when assessing the robustness of impacts for multiple measures across multiple subgroups of the study sample.

The Enhanced Reading Opportunities Study

Appendix Table E.3

Impacts on Reading Behaviors Composite Index,
Cohort 2 Respondent Sample and Subgroups

| Subgroup | Estimated Impact | P-Value for Estimated Impact |
|---|---------------------|------------------------------------|
| <u>All Schools</u> | 0.04 | 0.250 |
| <u>Programs</u> | | |
| Reading Apprenticeship schools | 0.03 | 0.493 |
| Xtreme Reading schools | 0.05 | 0.323 |
| <i>Difference in impacts</i> | <i>-0.02</i> | <i>0.822</i> |
| <u>Baseline comprehension performance</u> | | |
| 2.0-3.0 years below grade level | 0.06 | 0.282 |
| 3.1-4.0 years below grade level | 0.00 | 0.887 |
| 4.1-5.0 years below grade level | 0.05 | 0.318 |
| <i>Difference in impacts, 2.0-3.0 minus 3.1-4.0</i> | <i>0.06</i> | <i>0.533</i> |
| <i>Difference in impacts, 2.0-3.0 minus 4.1-5.0</i> | <i>0.01</i> | <i>0.912</i> |
| <u>Overage for grade^a</u> | | |
| Student is overage for grade | 0.03 | 0.654 |
| Student is not overage for grade | 0.04 | 0.262 |
| <i>Difference in impacts</i> | <i>-0.01</i> | <i>0.893</i> |
| <u>Language spoken at home</u> | | |
| Students from multilingual families | 0.03 | 0.488 |
| Students from English-only families | 0.04 | 0.387 |
| <i>Difference in impacts</i> | <i>-0.01</i> | <i>0.936</i> |
| <u>Teacher experience with the ERO program</u> | | |
| Teachers having taught two full years | 0.03 | 0.465 |
| Teachers having taught less than two full years | 0.06 | 0.324 |
| <i>Difference in impacts</i> | <i>-0.03</i> | <i>0.623</i> |
| Teachers having taught two full years - Year 1 | 0.03 | 0.407 |
| Teachers having taught two full years - Year 2 | 0.03 | 0.465 |
| <i>Difference in impacts</i> | <i>0.00</i> | <i>0.965</i> |
| <u>Second-year implementation strength</u> | | |
| Stronger implementation schools ^b | 0.03 | 0.561 |
| Weaker implementation schools ^c | 0.04 | 0.332 |
| <i>Difference in impacts</i> | <i>-0.01</i> | <i>0.896</i> |

(continued)

Appendix Table E.3 (continued)

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

NOTES: The reading behaviors composite index is the average of the standardized values of the three reading behavior measures: amount of school-related reading, amount of non-school-related reading, and use of reflective reading strategies. The values were standardized using the non-ERO group mean and standard deviation.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school and for random differences between the ERO and non-ERO groups in their baseline reading comprehension test scores and age at random assignment.

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

Rounding may cause slight discrepancies in calculating differences.

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

^bThe ERO programs in these schools were deemed to have reached an implementation level that was very well aligned to both the classroom learning environment and comprehension instruction dimensions of the program model by the spring site visit, and their ERO program began operating within 2 weeks of the start of the school year.

^cThe implementation fidelity of the ERO programs in these schools was deemed to be less than very well aligned to the classroom learning environment and/or comprehension instruction dimensions of the program model by the spring site visit, and/or their ERO program began operating 2 weeks or more after the start of the school year.

Appendix F

Impact Estimates Weighted for Nonresponse

As discussed in Appendix B, the two main data sources for this second-year impact analysis of the Enhanced Reading Opportunities (ERO) study are the student follow-up survey and the Group Reading Assessment and Diagnostic Examination (GRADE) assessment of student reading skills. The response analysis revealed several differences between students who completed the follow-up test and those who did not. Most notably, there were differences in response rates between the ERO group and the non-ERO group. In addition, nonrespondents were more likely to be overage for the ninth grade and to have lower pretest scores. As a result, students with these characteristics are underrepresented in the sample used to estimate impacts. The over- or underrepresentation of students with certain characteristics in the impact analysis sample may lead to findings that cannot be generalized to the original sample.

This appendix assesses the sensitivity of the impact estimates to the over- or underrepresentation of key baseline characteristics in the impact analysis sample. Specifically, it examines impact estimates that are weighted to account for differential response rates between the ERO and non-ERO groups and across high schools, as well as differential response rates associated with being overage for grade and baseline test scores. Sampling weights were constructed using multiple regressions in which response rates were predicted based on a student's baseline test score and an indicator of whether the student was overage for the ninth grade. Separate regressions were estimated for each high school and for the ERO students and non-ERO students within each school. The sampling weights were constructed as the inverse of the predicted response rate for each student in the full study sample.

These sampling weights ensure that each high school and the ERO and non-ERO groups within each high school can be represented in the impact analysis in the same proportion as they are in the full study sample. They also ensure that the distribution of overage-for-grade and baseline tests scores in the impact sample is equivalent to their representation in the full sample.

Appendix Table F.1 displays the weighted impact estimates for reading achievement for all 34 high schools and for the schools using each of the two supplemental reading programs. It shows that, together, the ERO programs produced a statistically significant weighted impact on reading comprehension of 0.9 standard score point (p -value = 0.035). This is slightly larger than the estimated impact for the respondent sample presented in Chapter 5 (0.8 standard score point; p -value = 0.042). The weighted impact of the Reading Apprenticeship Academic Literacy (RAAL) program on reading comprehension test scores is 1.4 standard score points and statistically significant (p -value = 0.035); this weighted estimate is of the same magnitude as the estimated impact for the respondent sample. The weighted impact of the Xtreme Reading program

The Enhanced Reading Opportunities Study

Appendix Table F.1

Impacts on Reading Achievement Weighted by School Response Rate,
Cohort 2 Follow-Up Respondent Sample

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|--|-------|------------------|---------------------|------------------------------------|------------------------------------|
| <u>All schools</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 89.9 | 89.1 | 0.9 * | 0.08 * | 0.035 |
| <i>Corresponding grade equivalent</i> | 6.1 | 5.9 | | | |
| <i>Corresponding percentile</i> | 24 | 23 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 93.3 | 93.4 | -0.1 | -0.01 | 0.818 |
| <i>Corresponding grade equivalent</i> | 7.7 | 7.7 | | | |
| <i>Corresponding percentile</i> | 32 | 32 | | | |
| Sample size | 1,264 | 907 | | | |
| <u>Reading Apprenticeship schools</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 90.0 | 88.6 | 1.4 * | 0.13 * | 0.013 |
| <i>Corresponding grade equivalent</i> | 6.1 | 5.8 | | | |
| <i>Corresponding percentile</i> | 25 | 22 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 93.2 | 93.7 | -0.5 | -0.05 | 0.357 |
| <i>Corresponding grade equivalent</i> | 7.7 | 7.8 | | | |
| <i>Corresponding percentile</i> | 31 | 32 | | | |
| Sample size | 645 | 470 | | | |
| <u>Xtreme Reading schools</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 89.8 | 89.5 | 0.3 | 0.03 | 0.630 |
| <i>Corresponding grade equivalent</i> | 6.1 | 6.0 | | | |
| <i>Corresponding percentile</i> | 24 | 24 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 93.3 | 93.0 | 0.3 | 0.03 | 0.582 |
| <i>Corresponding grade equivalent</i> | 7.7 | 7.7 | | | |
| <i>Corresponding percentile</i> | 32 | 31 | | | |
| Sample size | 619 | 437 | | | |

(continued)

Appendix Table F.1 (continued)

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

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

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

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

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (reading comprehension = 11.294; reading vocabulary = 11.099).

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.

on reading comprehension is not statistically significant (0.3 standard score point; p-value = 0.630), although it is slightly larger in magnitude than the unweighted impact estimate (0.2 standard score point; p-value = 0.672). Appendix Table F.1 also shows that the ERO programs did not have a statistically significant weighted impact on vocabulary test scores.

Appendix Table F.2 displays the weighted impacts on the reading behavior measures. These results are similar to those estimated for the respondent sample (see Tables 5.3 and 5.4).

In summary, differences between students who completed the follow-up test and survey and those who did not do not appear to change the underlying pattern of impacts on test scores or reading behaviors.

The Enhanced Reading Opportunities Study

Appendix Table F.2

Impacts on Reading Behaviors Weighted by School Response Rate,
Cohort 2 Follow-Up Respondent Sample

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|--|-------|---------------|------------------|------------------------------|------------------------------|
| <u>All schools</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 46.56 | 46.39 | 0.17 | 0.00 | 0.919 |
| Amount of non-school-related reading (prior month occurrences) | 30.20 | 29.80 | 0.40 | 0.01 | 0.781 |
| Use of reflective reading strategies (4-point scale) | 2.77 | 2.71 | 0.06 * | 0.09 * | 0.023 |
| Sample size | 1,260 | 901 | | | |
| <u>Reading Apprenticeship schools</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 50.59 | 49.37 | 1.22 | 0.03 | 0.616 |
| Amount of non-school-related reading (prior month occurrences) | 31.46 | 32.14 | -0.67 | -0.02 | 0.746 |
| Use of reflective reading strategies (4-point scale) | 2.78 | 2.74 | 0.04 | 0.07 | 0.208 |
| Sample size | 642 | 466 | | | |
| <u>Xtreme Reading schools</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 42.35 | 43.21 | -0.86 | -0.02 | 0.691 |
| Amount of non-school-related reading (prior month occurrences) | 28.88 | 27.23 | 1.65 | 0.04 | 0.401 |
| Use of reflective reading strategies (4-point scale) | 2.75 | 2.67 | 0.07 * | 0.11 * | 0.050 |
| Sample size | 618 | 435 | | | |

(continued)

Appendix Table F.2 (continued)

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

NOTES: The student follow-up survey was administered in spring 2007 at the end of students' ninth-grade year.

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

The estimated impact effect size is calculated as a proportion of the weighted standard deviation of the non-ERO group average (school-related reading standard deviation = 43.462; non-school related reading standard deviation = 37.334; use of reflective reading strategies standard deviation = 0.666).

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 3 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

Appendix G

**Impacts on Supplementary Measures of
Reading Achievement and Behaviors**

In an effort to understand more about the extent and nature of the impacts on student outcomes of the two supplemental literacy programs used in the Enhanced Reading Opportunities (ERO) study — Reading Apprenticeship Academic Literacy (RAAL) and Xtreme Reading — the ERO study team performed secondary impact analyses. These analyses fall into two categories. First, the supplemental analyses explore additional measures from the ERO follow-up student survey. These measures were created to complement the reading behaviors measures discussed in the report. They contribute to a more detailed picture of how the program changed or did not change students’ attitudes toward reading and their behavior in school. Second, the study team analyzed the impact of the ERO program on the percentage of students who were less than two years behind grade level in reading by the end of the school year. Given that students needed to be at least two years below grade level in reading to be eligible for the program, those students who have attained reading levels above this cutoff have succeeded in moving beyond the scope of the program during the school year.

Impacts on Students’ Attitudes and Perceptions of Reading and School

As discussed in Appendix A, the ERO follow-up student survey included a variety of questions related to students’ attitudes and perceptions of reading and school. Beyond the three reading behaviors measures discussed in the report, two other secondary measures were explored, including reading to learn and reading for enjoyment.

These measures are not included in the report because they were less directly related to ERO program goals or less likely to display short-term impacts. Appendix Table G.1 shows the impact findings for each of these measures. As shown in this table, Reading Apprenticeship Academic Literacy (RAAL) had a statistically significant impact on students’ enjoyment of reading (effect size of 0.13 standard deviation).

Impacts on the Percentage of Students No Longer Eligible for the ERO Programs

Both RAAL and Xtreme Reading attempt to accelerate literacy learning through their instructional programs to help struggling students attain the reading skill levels needed to succeed in high school classes. One way of measuring the impact of the ERO programs is to look at whether more ERO students are bridging this gap in skills during their first year of high school students who did not participate in ERO. To answer this question, the study team analyzed the programs’ impact on the percentage of students who were less than two years behind grade level in reading comprehension by the end of the school year, and, therefore, were no longer eligible for the program. The percentage of ERO students whose score on the follow-up

The Enhanced Reading Opportunities Study

Appendix Table G.1

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

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect | P-Value for Estimated Impact |
|--|-------|------------------|---------------------|-------------------------------|------------------------------------|
| <u>All schools</u> | | | | | |
| Reading to learn (4-point scale) | 2.73 | 2.69 | 0.04 | 0.06 | 0.117 |
| Reading for enjoyment (4-point scale) | 2.24 | 2.19 | 0.05 | 0.07 | 0.108 |
| Sample size | 1,260 | 901 | | | |
| <u>Reading Apprenticeship schools</u> | | | | | |
| Reading to learn (4-point scale) | 2.77 | 2.73 | 0.03 | 0.06 | 0.294 |
| Reading for enjoyment (4-point scale) | 2.32 | 2.22 | 0.10 * | 0.13 * | 0.026 |
| Sample size | 642 | 466 | | | |
| <u>Xtreme Reading schools</u> | | | | | |
| Reading to learn (4-point scale) | 2.69 | 2.65 | 0.04 | 0.07 | 0.240 |
| Reading for enjoyment (4-point scale) | 2.17 | 2.16 | 0.00 | 0.00 | 0.960 |
| Sample size | 618 | 435 | | | |

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

NOTES: The student follow-up survey was administered in spring 2007 at the end of students' ninth-grade year.

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

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (reading to learn standard deviation = 0.591; reading for enjoyment standard deviation = 0.776).

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 1.5 percent than of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

GRADE reading comprehension assessment was 98 or above and whose corresponding grade equivalent was greater than 7.9 were compared with the percentage of non-ERO students who scored at or above this level on the GRADE follow-up test.¹ As shown in Appendix Table G.2, the ERO program impacts for the entire sample and for each of the programs are not statistically significant at the 5 percent level. This indicates that a similar percentage of ERO and non-ERO students reached the benchmark level of reading achievement that would make them ineligible to reenroll in the ERO program.

¹A student who is two years below grade level at the end of ninth grade — and who would therefore still be eligible for the ERO program — reads at a grade equivalent (GE) of 7.9 (= 9.9 – 2.0). Hence, in order to no longer be eligible for the program, a student must read at a grade level of at least 8.0 at the end of ninth grade. A GE of 8.0 does not map perfectly onto GRADE standard scores, however. A standard score of 97 translates into a grade equivalent of 7.8, while a score of 98 maps onto a grade equivalent of 8.2. Since a score of 97 would render a student eligible for the program, a standard score of 98 and above is used as the cutoff for defining a student as no longer eligible for the ERO program.

The Enhanced Reading Opportunities Study

Appendix Table G.2

**Impacts on Percentage of Students No Longer Eligible for Program,
Cohort 2 Follow-Up Respondent Sample**

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|---|-------|------------------|---------------------|------------------------------------|------------------------------------|
| <u>All schools</u> | | | | | |
| No longer eligible for program ^a (%) | 22.78 | 19.90 | 2.88 | 0.07 | 0.093 |
| Sample size | 1,264 | 907 | | | |
| <u>Reading Apprenticeship schools</u> | | | | | |
| No longer eligible for program (%) | 22.79 | 18.78 | 4.01 | 0.10 | 0.090 |
| Sample size | 645 | 470 | | | |
| <u>Xtreme Reading schools</u> | | | | | |
| No longer eligible for program (%) | 22.78 | 21.05 | 1.73 | 0.04 | 0.487 |
| Sample size | 619 | 437 | | | |

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

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

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

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (standard deviation = 40.317).

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

Rounding may cause slight discrepancies in calculating sums and differences.

^aStudents with scores on the GRADE pretest between 2 and 5 years below grade level were eligible for the program. Students are considered no longer eligible for the program if their score on the follow-up GRADE assessment corresponded to a grade equivalent greater than 7.9 (standard score of 98), suggesting that the student is now less than two years behind grade level.

Appendix H

Impacts for Student Subgroups

While all students in the study sample for the Enhanced Reading Opportunities (ERO) evaluation had baseline reading comprehension skills that were two to five years below grade level in the spring of eighth grade, the ERO study sample includes a diverse population of students. With this diversity in mind, the ERO evaluation was designed to allow for the estimation of impacts for key subgroups of students who face especially challenging barriers to literacy development and overall school performance in high school. For example, prior research has shown that especially low literacy levels, evidence of failure in prior grades, and having English as a second language are powerful predictors of school success.¹

This appendix examines variation in ERO program impacts for subgroups of students defined by their baseline reading comprehension test scores, whether they were overage for the ninth grade, and whether a language other than English was spoken in their homes. Among the respondent sample, 39 percent had baseline test scores that indicate reading levels that were four to five years below grade level at the times of testing, and another 29 percent were reading from three to four years below grade level. Also, 26 percent of the students in the study sample were overage for the ninth grade, which is used to indicate that a student was retained in a prior grade. Half of the students in the sample lived in households where a language other than English was spoken.

- **Differences in impacts across subgroups of students with different baseline reading comprehension test scores are not statistically significant for reading comprehension and reading behaviors, but are statistically significant for reading vocabulary.**

Appendix Tables H.1 and H.2 correspond to the top panel of Table 5.5 and present impact findings for the subgroups of students defined by their baseline reading comprehension test scores. Appendix Table H.1 indicates that the ERO programs did not produce statistically significant impacts on reading comprehension or reading vocabulary test scores for any of the three subgroups defined by baseline test scores.

That said, the bottom section of Appendix Table H.1 indicates that the impact of the programs on reading vocabulary scores differs by a statistically significant amount across some of the subgroups. Specifically, the estimated impact on reading vocabulary scores for students who were two to three years below grade level at baseline is significantly larger than the estimated impact for students who were four to five years below grade level at baseline (difference in effect size = 0.23 standard deviation; p-value = 0.019).

¹Roderick (1993); Fine (1988).

The Enhanced Reading Opportunities Study

Appendix Table H.1

Impacts on Reading Achievement,
Cohort 2 Follow-Up Respondent Sample,
by Baseline Reading Comprehension Performance

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|---|------|---------------|------------------|------------------------------|------------------------------|
| <u>2.0-3.0 years below grade level</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 94.3 | 93.4 | 0.9 | 0.09 | 0.170 |
| <i>Corresponding grade equivalent</i> | 7.2 | 7.0 | | | |
| <i>Corresponding percentile</i> | 34 | 33 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 97.7 | 96.5 | 1.2 | 0.12 | 0.088 |
| <i>Corresponding grade equivalent</i> | 8.5 | 8.2 | | | |
| <i>Corresponding percentile</i> | 42 | 39 | | | |
| Sample size | 404 | 301 | | | |
| <u>3.1-4.0 years below grade level</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 91.1 | 90.2 | 0.9 | 0.09 | 0.273 |
| <i>Corresponding grade equivalent</i> | 6.3 | 6.1 | | | |
| <i>Corresponding percentile</i> | 26 | 25 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 94.3 | 93.9 | 0.4 | 0.04 | 0.625 |
| <i>Corresponding grade equivalent</i> | 7.8 | 7.8 | | | |
| <i>Corresponding percentile</i> | 34 | 33 | | | |
| Sample size | 350 | 269 | | | |
| <u>4.1-5.0 years below grade level</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 86.1 | 85.3 | 0.8 | 0.08 | 0.210 |
| <i>Corresponding grade equivalent</i> | 5.1 | 5.0 | | | |
| <i>Corresponding percentile</i> | 17 | 15 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 89.5 | 90.6 | -1.1 | -0.11 | 0.106 |
| <i>Corresponding grade equivalent</i> | 7.1 | 7.3 | | | |
| <i>Corresponding percentile</i> | 23 | 26 | | | |
| Sample size | 510 | 337 | | | |

(continued)

Appendix Table H.1 (continued)

| Difference in Impacts Between Subgroups | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
|---|--------------------------|---|---------------------------|
| <u>2.0-3.0 minus 3.1-4.0</u> | | | |
| Reading comprehension standard score | 0.0 | 0.00 | 0.968 |
| Reading vocabulary standard score | 0.8 | 0.09 | 0.414 |
| <u>2.0-3.0 minus 4.1-5.0</u> | | | |
| Reading comprehension standard score | 0.1 | 0.01 | 0.897 |
| Reading vocabulary standard score | 2.3 * | 0.23 * | 0.019 |

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

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

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

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

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

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

Rounding may cause slight discrepancies in calculating sums and differences.

The Enhanced Reading Opportunities Study
Appendix Table H.2
Impacts on Reading Behaviors,
Cohort 2 Follow-Up Respondent Sample,
by Baseline Reading Comprehension Performance

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|---|------|------------------|---------------------|------------------------------------|------------------------------------|
| <u>2.0-3.0 years behind grade level</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 47.1 | 47.3 | -0.3 | -0.01 | 0.925 |
| Amount of non-school-related reading (prior month occurrences) | 28.7 | 26.4 | 2.3 | 0.07 | 0.353 |
| Use of reflective reading strategies (4-point scale) | 2.8 | 2.7 | 0.1 | 0.13 | 0.070 |
| Sample size | 403 | 299 | | | |
| <u>3.1-4.0 years behind grade level</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 48.0 | 46.8 | 1.1 | 0.03 | 0.711 |
| Amount of non-school-related reading (prior month occurrences) | 30.4 | 31.9 | -1.4 | -0.04 | 0.595 |
| Use of reflective reading strategies (4-point scale) | 2.8 | 2.8 | 0.0 | 0.05 | 0.552 |
| Sample size | 347 | 268 | | | |
| <u>4.1-5.0 years behind grade level</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 44.5 | 44.3 | 0.3 | 0.01 | 0.920 |
| Amount of non-school-related reading (prior month occurrences) | 30.4 | 29.2 | 1.3 | 0.04 | 0.582 |
| Use of reflective reading strategies (4-point scale) | 2.7 | 2.6 | 0.1 | 0.12 | 0.119 |
| Sample size | 510 | 334 | | | |

(continued)

Appendix Table H.2 (continued)

| Difference in Impacts Between Subgroups | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
|---|--------------------------|---|---------------------------|
| <u>2.0-3.0 minus 3.1-4.0</u> | | | |
| Amount of school-related reading | -1.4 | -0.04 | 0.739 |
| Amount of non-school-related reading | 3.7 | 0.11 | 0.307 |
| Use of reflective reading strategies | 0.0 | 0.08 | 0.445 |
| <u>2.0-3.0 minus 4.1-5.0</u> | | | |
| Amount of school-related reading | -0.5 | -0.01 | 0.891 |
| Amount of non-school-related reading | 1.1 | 0.03 | 0.755 |
| Use of reflective reading strategies | 0.0 | 0.01 | 0.924 |

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

NOTES: The student follow-up survey was administered in spring 2007 at the end of students' ninth-grade year.

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

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 38.322; non-school-related reading standard deviation = 32.976; use of reflective reading strategies standard deviation = 0.592).

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

For each of the above measures, data are missing for no more than 3.4 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

Appendix Table H.2 shows that the ERO programs had a statistically significant impact on the use of reflective reading strategies for students whose scores were from four to five years below grade level. However, it cannot be concluded that the impacts of the ERO programs on this measure was different for students with different baseline reading comprehension test scores for two reasons. First, the qualifying tests conducted for this subgroup of students (see Appendix E) indicate that the ERO programs did not produce a statistically significant impact on the composite index of the three reading behavior measures. Second, the difference between the impact on reading strategies for this subgroup and the impact for each of the other two subgroups is not statistically significant.

- **Differences in impacts across subgroups of students who were overage for the ninth grade or not overage for the ninth grade are not statistically significant.**

Appendix Tables H.3 and H.4 correspond to the middle panel of Table 5.5 and present impact findings for the subgroups of students defined by whether they were overage for the ninth grade. Appendix Table H.3 indicates that the ERO programs produced positive and statistically significant impacts on reading comprehension test scores for students who were not overage for grade (effect size = 0.10 standard deviation; p-value = 0.028). Although the impact on reading comprehension test scores for this group is statistically significant, the difference between this impact and the impact for students who were overage for grade is not statistically significant. Appendix Table H.4 shows that the ERO programs also produced a statistically significant impact on the use of reflective reading strategies for students who were not overage for grade (effect size = 0.10 standard deviation; p-value = 0.034). However, the difference between this impact and the impact for students who were overage for grade is not statistically significant. Thus, it cannot be concluded that the impacts of the ERO programs on these measures were different for students who were not overage for grade compared to those who were. Moreover, the qualifying tests conducted for students who are not overage for grade (see Appendix E) indicate that the ERO programs did not produce a statistically significant impact for this subgroup on the composite index of the three reading behavior measures.

- **Differences in impacts across subgroups of students from multilingual families and those from English-only families are not statistically significant.**

Appendix Tables H.5 and H.6 correspond to the bottom panel of Table 5.5 and present impact findings for the subgroups of students defined by whether a language other than English was spoken in their homes. Appendix Table H.5 indicates that the ERO programs did not produce statistically significant impacts on reading comprehension test scores for either of the two

subgroups defined by language spoken at home. Similarly, Appendix Table H.6 indicates that the ERO programs did not have a statistically significant impact on reading behaviors for students from multilingual families or from English-only families.

An overall F-test was used to test whether there is systematic variation in impacts on reading comprehension across the various subgroups defined by baseline reading comprehension test scores, overage status, and language spoken at home.² This test indicates that the overall variation in impacts across all of these subgroups is not statistically significant (F-statistic = 1.242; p-value = 0.291), further suggesting that the statistical significance of reading comprehension impacts for specific subgroups should be interpreted with caution.³

²This test was conducted by adding a set of interactions between the treatment indicator and the subgroup indicators (two subgroup indicators for baseline reading comprehension score, an indicator for overage status, and an indicator for language spoken at home) to the impact regression model fit to the Cohort 2 follow-up respondent sample. An F-test was then used to test for the joint significance of these interaction terms.

³This overall F-test was also conducted for reading vocabulary, the three reading behaviors outcomes, and the reading behavior composite. These tests indicate that there is overall variation in impacts across the student subgroups for reading vocabulary (F-statistic = 3.69; p-value = 0.005), which is partially driven by the difference in impacts across subgroups defined by reading comprehension scores at baseline. There is also overall variation in impacts on the frequency of school-related reading (F-statistic = 2.70; p-value = 0.029), and the reading behaviors composite (F-statistic = 2.64; p-value = 0.032).

The Enhanced Reading Opportunities Study
Appendix Table H.3
Impacts on Reading Achievement,
Cohort 2 Follow-Up Respondent 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 |
|--|-----------|---------------|-----------------------|-----------------------------------|------------------------------|
| <u>Overage for grade^a</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 87.4 | 87.0 | 0.5 | 0.05 | 0.592 |
| <i>Corresponding grade equivalent</i> | 5.4 | 5.3 | | | |
| <i>Corresponding percentile</i> | 19 | 18 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 91.3 | 91.2 | 0.1 | 0.01 | 0.939 |
| <i>Corresponding grade equivalent</i> | 7.5 | 7.4 | | | |
| <i>Corresponding percentile</i> | 27 | 27 | | | |
| Sample size | 323 | 245 | | | |
| <u>Not overage for grade</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 91.0 | 90.0 | 1.0 * | 0.10 * | 0.028 |
| <i>Corresponding grade equivalent</i> | 6.3 | 6.1 | | | |
| <i>Corresponding percentile</i> | 26 | 25 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 94.2 | 94.2 | 0.0 | 0.00 | 0.996 |
| <i>Corresponding grade equivalent</i> | 7.8 | 7.8 | | | |
| <i>Corresponding percentile</i> | 33 | 33 | | | |
| Sample size | 941 | 662 | | | |
| Difference in Impacts Between Subgroups | | | | | |
| | | | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
| <u>Overage minus not overage</u> | | | | | |
| Reading comprehension standard score | | | -0.5 | -0.05 | 0.587 |
| Reading vocabulary standard score | | | 0.1 | 0.01 | 0.949 |

(continued)

Appendix Table H.3 (continued)

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

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

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

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

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

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

Rounding may cause slight discrepancies in calculating sums and differences.

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

The Enhanced Reading Opportunities Study
Appendix Table H.4
Impacts on Reading Behaviors,
Cohort 2 Follow-Up Respondent 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 |
|--|-----------|---------------|-----------------------|-----------------------------------|------------------------------|
| <u>Overage for grade^a</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 45.4 | 44.7 | 0.7 | 0.02 | 0.841 |
| Amount of non-school-related reading (prior month occurrences) | 31.7 | 31.5 | 0.2 | 0.01 | 0.940 |
| Use of reflective reading strategies (4-point scale) | 2.8 | 2.7 | 0.0 | 0.05 | 0.585 |
| Sample size | 320 | 240 | | | |
| <u>Not overage for grade</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 46.6 | 46.3 | 0.2 | 0.01 | 0.891 |
| Amount of non-school-related reading (prior month occurrences) | 29.3 | 28.4 | 0.8 | 0.03 | 0.598 |
| Use of reflective reading strategies (4-point scale) | 2.8 | 2.7 | 0.1 * | 0.10 * | 0.034 |
| Sample size | 940 | 661 | | | |
| | | | | | |
| Difference in Impacts Between Subgroups | | | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
| <u>Overage minus not overage</u> | | | | | |
| Amount of school-related reading | | | 0.4 | 0.01 | 0.910 |
| Amount of non-school-related reading | | | -0.6 | -0.02 | 0.863 |
| Use of reflective reading strategies | | | 0.0 | -0.06 | 0.574 |

(continued)

Appendix Table H.4 (continued)

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

NOTES: The student follow-up survey was administered in spring 2007 at the end of students' ninth-grade year.

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

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 38.322; non-school-related reading standard deviation = 32.976; use of reflective reading strategies standard deviation = 0.592).

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

For each of the above measures, data are missing for no more than 4.2 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

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

The Enhanced Reading Opportunities Study

Appendix Table H.5

Impacts on Reading Achievement,
Cohort 2 Follow-Up Respondent Sample,
by Language Spoken at Home

| Outcome | ERO Group | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|---|-----------|---------------|-----------------------|-----------------------------------|------------------------------|
| <u>Students from multilingual families</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 89.6 | 88.8 | 0.8 | 0.08 | 0.181 |
| <i>Corresponding grade equivalent</i> | 6.0 | 5.8 | | | |
| <i>Corresponding percentile</i> | 24 | 22 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 92.0 | 92.2 | -0.2 | -0.02 | 0.768 |
| <i>Corresponding grade equivalent</i> | 7.6 | 7.6 | | | |
| <i>Corresponding percentile</i> | 29 | 29 | | | |
| Sample size | 621 | 458 | | | |
| <u>Students from English-only families</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 90.6 | 89.7 | 0.9 | 0.09 | 0.107 |
| <i>Corresponding grade equivalent</i> | 6.2 | 6.0 | | | |
| <i>Corresponding percentile</i> | 25 | 24 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 94.9 | 94.5 | 0.3 | 0.03 | 0.550 |
| <i>Corresponding grade equivalent</i> | 7.9 | 7.9 | | | |
| <i>Corresponding percentile</i> | 35 | 34 | | | |
| Sample size | 643 | 449 | | | |
| Difference in Impacts Between Subgroups | | | | | |
| | | | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
| <u>Multilingual minus English-only</u> | | | | | |
| Reading comprehension standard score | | | -0.1 | -0.01 | 0.868 |
| Reading vocabulary standard score | | | -0.5 | -0.05 | 0.531 |

(continued)

Appendix Table H.5 (continued)

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

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

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

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

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

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

Rounding may cause slight discrepancies in calculating sums and differences.

The Enhanced Reading Opportunities Study
Appendix Table H.6
Impacts on Reading Behaviors,
Cohort 2 Follow-Up Respondent Sample,
by Language Spoken at Home

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect | P-Value for Estimated Impact |
|---|------|------------------|--------------------------|---|------------------------------------|
| <u>Students from multilingual families</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 47.4 | 45.9 | 1.5 | 0.04 | 0.532 |
| Amount of non-school-related reading (prior month occurrences) | 30.3 | 30.8 | -0.6 | -0.02 | 0.778 |
| Use of reflective reading strategies (4-point scale) | 2.8 | 2.7 | 0.0 | 0.07 | 0.224 |
| Sample size | 621 | 455 | | | |
| <u>Students from English-only families</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 45.2 | 46.1 | -0.8 | -0.02 | 0.709 |
| Amount of non-school-related reading (prior month occurrences) | 29.5 | 28.0 | 1.4 | 0.04 | 0.458 |
| Use of reflective reading strategies (4-point scale) | 2.8 | 2.7 | 0.1 | 0.10 | 0.090 |
| Sample size | 639 | 446 | | | |
| | | | | | |
| Difference in Impacts Between Subgroups | | | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
| <u>Multilingual minus English-only</u> | | | | | |
| Amount of school-related reading | | | 2.3 | 0.06 | 0.476 |
| Amount of non-school-related reading | | | -2.0 | -0.06 | 0.473 |
| Use of reflective reading strategies | | | 0.0 | -0.03 | 0.765 |
| | | | | | |

(continued)

Appendix Table H.6 (continued)

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

NOTES: The student follow-up survey was administered in spring 2007 at the end of students' ninth-grade year.

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

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 38.322; non-school-related reading standard deviation = 32.976; use of reflective reading strategies standard deviation = 0.592).

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

For each of the above measures, data are missing for no more than 3.4 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

Appendix I

**The Relationship Between Impacts and
Second-Year Implementation**

This appendix presents results from exploratory analyses that examine the relationship between school-level impacts and various aspects of implementation in the second year of the Enhanced Reading Opportunities (ERO) study. As noted in Chapter 5, school-to-school variation in impacts was not statistically significant in the second year of the study; hence, these exploratory analyses are unlikely to provide conclusive information about the factors that predict differences in impacts across schools. Nonetheless, the relationship between impacts and certain policy-relevant aspects of implementation was examined and the results of these analyses are presented in this appendix.

Given that the supplemental literacy programs used in the ERO study — Reading Apprenticeship Academic Literacy (RAAL) and Xtreme Reading — operated in more favorable conditions in the second year of the study than in the first (as discussed in Chapter 3), this exploratory analysis focuses on the relationship between school-level impacts and whether a school’s ERO program was particularly exemplary in terms of its implementation. This stands in contrast to the exploratory analysis conducted in the first-year report, which focuses instead on the relationship between school-level impacts and program implementation that was especially problematic.

The first set of results in this appendix examines the relationship between impacts and the experience of a school’s ERO teacher with the program. Specifically, this analysis investigates whether the impact of the programs in the second year of the study differs between the 25 schools whose ERO teacher had also taught the entire first year, and the 9 schools whose ERO teacher was new to the program in the second year of the study or had been brought in as a replacement midway through the first year. In addition, the analysis compares the impacts produced by the 25 returning teachers in their second year of teaching the program (Cohort 2) and the impacts produced in their first year of teaching the program (Cohort 1).

The second set of results in this appendix examines the relationship between impacts and the strength of program implementation. Two types of relationship are explored: (1) the relationship between impacts and implementation fidelity and (2) the relationship between impacts and the number of weeks between the start of the school year and program start-up. The indicators are then combined, and impacts are then estimated for two groups of sites defined by whether the implementation of their ERO program was strong on both of these indicators.

It is important to note that the analyses presented in this appendix are exploratory in nature and cannot be used to establish causal links between these aspects of implementation and variation in program impacts across the sites.

Variation in Impacts Across Schools

Appendix Table I.1 is the counterpart to Figure 5.2. It presents the reading comprehension impact estimates of each of the 34 participating high schools in ascending order. It also includes the standard error and 95 percent confidence intervals for these impacts. One of the 34 schools had a statistically significant positive impact. A composite F-test was used to assess whether the school-level impacts on reading comprehension test scores are statistically equivalent. The F-value is 0.91, and the p-value is 0.614, indicating that the school-to-school variation in impacts may have occurred by chance.

Relationship Between Impacts and Teacher Experience with the ERO Program

Impacts by ERO Teacher's Experience in the Second Year

Appendix Tables I.2 and I.3 correspond with the upper half of the top panel of Table 5.6. These tables present impacts on reading test scores and reading behaviors, consecutively, for the 25 schools whose ERO teacher in the second year of the study had also taught the entire first year of the study (that is, teachers having taught two full years of the program by the end of the study) and for the 9 schools whose ERO teacher was either new at the start of the second year or had replaced another teacher midway through the first year of the study (that is, teachers having taught less than two full years of the program by the end of the study).

Appendix Table I.2 indicates that while the impact of the ERO programs was positive and statistically significant in sites where the ERO teacher had greater experience with the program (effect size = 0.09 standard deviation; p-value = 0.050), this impact is not statistically different from the impact for less experienced teachers (effect size = 0.06 standard deviation; p-value = 0.487). Hence, it cannot be concluded that impacts were larger in sites with more experienced teachers.

Similarly, Appendix Table I.3 shows that the impact of the ERO programs on the use of reflective reading strategies was also positive and statistically significant in schools where the ERO teacher had more experience with the program (effect size = 0.10 standard deviation; p-value = 0.046); however, this impact is not statistically different from the impact of the ERO programs on reading strategies in schools where the ERO teacher was less experienced (effect size = 0.07 standard deviation; p-value = 0.441). Here, again, it cannot be concluded that impacts were larger in sites with more experienced teachers. In addition, the statistical significance of the impact among more experienced teachers is called into question by the qualifying tests in Appendix E, which shows that the impact of the programs on the reading behavior composite index in sites with a more experienced ERO teacher is not statistically significant.

The Enhanced Reading Opportunities Study

Appendix Table I.1

**Fixed-Effect Impact Estimates on Reading Comprehension,
by School**

| Variable | Impact Estimate | Standard Error | 95% Confidence Interval | |
|-----------------------|--------------------|-------------------|----------------------------|------|
| School 1 ^a | -3.7 | 2.74 | -9.11 | 1.65 |
| School 2 | -3.6 | 2.39 | -8.33 | 1.04 |
| School 3 | -2.5 | 2.37 | -7.19 | 2.11 |
| School 4 | -2.5 | 2.07 | -6.57 | 1.55 |
| School 5 | -1.9 | 2.48 | -6.73 | 3.00 |
| School 6 | -1.7 | 2.63 | -6.82 | 3.50 |
| School 7 | -1.3 | 1.83 | -4.88 | 2.29 |
| School 8 | -0.7 | 2.15 | -4.89 | 3.53 |
| School 9 | -0.4 | 2.13 | -4.61 | 3.74 |
| School 10 | 0.1 | 2.24 | -4.33 | 4.45 |
| School 11 | 0.4 | 2.44 | -4.39 | 5.20 |
| School 12 | 0.4 | 2.10 | -3.70 | 4.53 |
| School 13 | 0.5 | 2.00 | -3.46 | 4.37 |
| School 14 | 0.7 | 2.33 | -3.89 | 5.25 |
| School 15 | 0.8 | 2.48 | -4.06 | 5.68 |
| School 16 | 1.0 | 2.33 | -3.53 | 5.62 |
| School 17 | 1.1 | 2.18 | -3.17 | 5.37 |
| School 18 | 1.6 | 2.31 | -2.93 | 6.12 |
| School 19 | 1.6 | 2.21 | -2.71 | 5.94 |
| School 20 | 1.6 | 2.90 | -4.05 | 7.32 |
| School 21 | 1.7 | 2.11 | -2.44 | 5.85 |
| School 22 | 1.8 | 2.55 | -3.17 | 6.82 |
| School 23 | 2.1 | 2.22 | -2.31 | 6.42 |
| School 24 | 2.1 | 2.58 | -2.97 | 7.13 |
| School 25 | 2.4 | 2.40 | -2.33 | 7.07 |
| School 26 | 2.5 | 2.28 | -1.96 | 6.99 |
| School 27 | 2.7 | 2.21 | -1.62 | 7.03 |

(continued)

Appendix Table I.1 (continued)

| Variable | Impact Estimate | Standard Error | 95% Confidence Interval | |
|-----------|-----------------|----------------|-------------------------|-------|
| School 28 | 2.9 | 4.52 | -6.00 | 11.72 |
| School 29 | 3.3 | 2.66 | -1.96 | 8.48 |
| School 30 | 3.9 | 2.94 | -1.91 | 9.63 |
| School 31 | 4.0 | 2.59 | -1.03 | 9.13 |
| School 32 | 5.3 * | 2.13 | 1.09 | 9.43 |
| School 33 | 5.6 | 3.50 | -1.24 | 12.50 |
| School 34 | 6.2 | 3.89 | -1.43 | 13.81 |

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

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

The fixed-effect estimated impacts are the regression-adjusted coefficients on the interaction between school and treatment using ordinary least squares, controlling for blocking of random assignment by school and for random differences between the ERO and non-ERO groups in their baseline reading comprehension test scores and age at random assignment.

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

A composite F-test was used to assess whether the school-level impacts on reading comprehension test scores are statistically equivalent. The F-value is 0.91, and the p-value is 0.614, indicating that the school-to-school variation in impacts is likely to have occurred by chance.

^aThe schools are listed in ascending order by their impact estimate.

The Enhanced Reading Opportunities Study
Appendix Table I.2
Impacts on Reading Achievement,
Cohort 2 Follow-Up Respondent Sample,
by Teacher Experience with the ERO Program

| Outcome | ERO Group | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|---|-----------|---------------|-----------------------|-----------------------------------|------------------------------|
| <u>Teachers having taught two full years</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 90.0 | 89.1 | 0.9 * | 0.09 * | 0.050 |
| <i>Corresponding grade equivalent</i> | 6.1 | 5.9 | | | |
| <i>Corresponding percentile</i> | 25 | 23 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 93.1 | 93.3 | -0.1 | -0.01 | 0.791 |
| <i>Corresponding grade equivalent</i> | 7.7 | 7.7 | | | |
| <i>Corresponding percentile</i> | 31 | 32 | | | |
| Sample size | 942 | 680 | | | |
| <u>Teachers having taught less than two full years^a</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 90.4 | 89.8 | 0.6 | 0.06 | 0.487 |
| <i>Corresponding grade equivalent</i> | 6.2 | 6.1 | | | |
| <i>Corresponding percentile</i> | 25 | 24 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 94.4 | 94.1 | 0.3 | 0.03 | 0.694 |
| <i>Corresponding grade equivalent</i> | 7.8 | 7.8 | | | |
| <i>Corresponding percentile</i> | 34 | 33 | | | |
| Sample size | 322 | 227 | | | |
| | | | | | |
| Difference in Impacts Between Subgroups | | | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
| <u>Entire year minus new or less than a year</u> | | | | | |
| Reading comprehension standard score | | | 0.4 | 0.04 | 0.710 |
| Reading vocabulary standard score | | | -0.5 | -0.05 | 0.636 |

(continued)

Appendix Table I.2 (continued)

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

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

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

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

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

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

Rounding may cause slight discrepancies in calculating sums and differences.

^aIncludes both ERO teachers who started in Year 2, and replacement teachers who began mid-year in Year 1.

The Enhanced Reading Opportunities Study
Appendix Table I.3
Impacts on Reading Behaviors,
Cohort 2 Follow-Up Respondent Sample,
by Teacher Experience with the ERO Program

| Outcome | Non-ERO ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|---|----------------|------------------|--------------------------|---|------------------------------------|
| <u>Teachers having taught two full years</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 45.6 | 46.6 | -1.0 | -0.03 | 0.590 |
| Amount of non-school-related reading (prior month occurrences) | 29.2 | 28.8 | 0.5 | 0.01 | 0.774 |
| Use of reflective reading strategies (4-point scale) | 2.8 | 2.7 | 0.06 * | 0.10 * | 0.046 |
| Sample size | 941 | 678 | | | |
| <u>Teachers having taught less than two full years^a</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 48.3 | 43.9 | 4.4 | 0.11 | 0.150 |
| Amount of non-school-related reading (prior month occurrences) | 31.7 | 31.2 | 0.6 | 0.02 | 0.843 |
| Use of reflective reading strategies (4-point scale) | 2.8 | 2.7 | 0.04 | 0.07 | 0.441 |
| Sample size | 319 | 223 | | | |
| | | | | | |
| Difference in Impacts Between Subgroups | | | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
| <u>Entire year minus new or less than a year</u> | | | | | |
| Amount of school-related reading | | | -5.4 | -0.14 | 0.131 |
| Amount of non-school-related reading | | | -0.1 | 0.00 | 0.976 |
| Use of reflective reading strategies | | | 0.02 | 0.03 | 0.765 |
| | | | | | |

(continued)

Appendix Table I.3 (continued)

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

NOTES: The student follow-up survey was administered in spring 2007 at the end of students' ninth-grade year.

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

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 38.322; non-school-related reading standard deviation = 32.976; use of reflective reading strategies standard deviation = 0.592).

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

For each of the above measures, data are missing for no more than 3.2 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

^aIncludes both ERO teachers who started in Year 2, and replacement teachers who began mid-year in Year 1.

Impacts by Study Year for Teachers Having Taught Both Years

Appendix Tables I.4 and I.5 correspond with the lower half of the top panel of Table 5.6. These tables present impacts on reading test scores and reading behaviors, consecutively, for the 25 schools whose ERO teacher taught two full years of the program, during each of the two years that they taught the program.

Appendix Table I.4 indicates that the impact produced by returning teachers on reading comprehension scores was positive and statistically significant in both years (effect size = 0.11 standard deviation in the first year and 0.09 standard deviation in the second year). However, the difference between these two impacts is not statistically significant.

Appendix Table I.5 shows that returning ERO teachers produced a larger impact on students' use of reading strategies in their second year of teaching the program (effect size = 0.10 standard deviation; p-value = 0.046) than in their first year (effect size = -0.05; p-value = 0.331) and that the difference between these impacts is statistically significant. Given that the composite qualifying tests in Appendix E indicate that the difference in impacts between Year 1 and Year 2 on the composite measure of reading behaviors is not statistically significant, one cannot be certain that the statistical significance of this finding is not the result of chance.

Relationship Between Impacts and Program Implementation

Relationship Between Impacts and Implementation Fidelity

Appendix Tables I.6 and I.7 correspond with the second panel of Table 5.6. These tables present impacts on reading test scores and reading behaviors, consecutively, by implementation fidelity at the second-year spring site visit. Schools are categorized as having ERO programs that are either “very well aligned,” “well aligned,” “moderately aligned,” or “poorly aligned” with their respective program models (see Chapter 3 and Appendix D for the definition of these categories). Note that impacts are not presented for the “poorly aligned” category because only one site was deemed to have poorly aligned implementation fidelity at the spring site visit. In particular, one cannot make generalized statements about the impact produced by schools with “poorly aligned” programs based on only one school rated in that category. The results are also suppressed to protect the identity of this particular school and its associated impact.¹

¹Given that there was only one school whose implementation fidelity was poorly aligned to the program models, one possibility would have been to combine the “moderately aligned” and “poorly aligned” schools into one category. However, these two categories were kept separate in order to preserve the same fidelity groupings that were used in the first-year report of the ERO study (see Kemple et al. 2008), as well as in Chapter 3 of the present report.

The Enhanced Reading Opportunities Study

Appendix Table I.4

Impacts on Reading Achievement in
Schools Where Teacher Taught Two Full Years of the ERO Program,
by Cohort Respondent Sample

| Outcome | ERO Group | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|---------------------------------------|-----------|---------------|-----------------------|------------------------------|------------------------------|
| <u>Cohort 1</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 90.7 | 89.6 | 1.1 * | 0.11 * | 0.014 |
| <i>Corresponding grade equivalent</i> | 6.2 | 6.0 | | | |
| <i>Corresponding percentile</i> | 26 | 24 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 93.6 | 93.1 | 0.5 | 0.05 | 0.274 |
| <i>Corresponding grade equivalent</i> | 7.8 | 7.7 | | | |
| <i>Corresponding percentile</i> | 32 | 31 | | | |
| Sample size | 1040 | 748 | | | |
| <u>Cohort 2</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 90.0 | 89.1 | 0.9 * | 0.09 * | 0.050 |
| <i>Corresponding grade equivalent</i> | 6.1 | 5.9 | | | |
| <i>Corresponding percentile</i> | 25 | 23 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 93.1 | 93.3 | -0.1 | -0.01 | 0.791 |
| <i>Corresponding grade equivalent</i> | 7.7 | 7.7 | | | |
| <i>Corresponding percentile</i> | 31 | 32 | | | |
| Sample size | 942 | 680 | | | |
| <hr/> | | | | | |
| Difference in Impacts Between Cohorts | | | Difference in Impacts | Difference in Impact Sizes | P-Value for Difference |
| <u>Cohort 2 minus Cohort 1</u> | | | | | |
| Reading comprehension standard score | | | -0.2 | -0.02 | 0.782 |
| Reading vocabulary standard score | | | -0.6 | -0.06 | 0.341 |

(continued)

Appendix Table I.4 (continued)

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

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

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

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

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

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

Rounding may cause slight discrepancies in calculating sums and differences.

The Enhanced Reading Opportunities Study
Appendix Table I.5
Impacts on Reading Behaviors in
Schools Where Teacher Taught Two Full Years of the ERO Program,
by Cohort Respondent Sample

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|---|------|------------------|--------------------------|---|------------------------------------|
| <u>Cohort 1</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 45.3 | 43.3 | 2.0 | 0.05 | 0.328 |
| Amount of non-school-related reading (prior month occurrences) | 27.5 | 25.1 | 2.4 | 0.07 | 0.095 |
| Use of reflective reading strategies (4-point scale) | 2.61 | 2.64 | -0.03 | -0.05 | 0.331 |
| Sample size | 1039 | 747 | | | |
| <u>Cohort 2</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 45.6 | 46.6 | -1.0 | -0.03 | 0.590 |
| Amount of non-school-related reading (prior month occurrences) | 29.2 | 28.8 | 0.5 | 0.01 | 0.774 |
| Use of reflective reading strategies (4-point scale) | 2.76 | 2.70 | 0.06 * | 0.10 * | 0.046 |
| Sample size | 941 | 678 | | | |
| | | | | | |
| Difference in Impacts Between Subgroups | | | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
| <u>Cohort 2 minus Cohort 1</u> | | | | | |
| Amount of school-related reading | | | -3.1 | -0.08 | 0.276 |
| Amount of non-school-related reading | | | -2.0 | -0.06 | 0.366 |
| Use of reflective reading strategies | | | 0.09 * | 0.15 * | 0.038 |

(continued)

Appendix Table I.5 (continued)

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

NOTES: The student follow-up survey was administered in spring 2007 at the end of students' ninth-grade year.

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

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 38.322; non-school-related reading standard deviation = 32.976; use of reflective reading strategies standard deviation = 0.592).

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

For each of the above measures, data are missing for no more than 3.6 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

The Enhanced Reading Opportunities Study
Appendix Table I.6
Impacts on Reading Achievement,
Cohort 2 Follow-Up Respondent Sample,
by Program Implementation Fidelity at Spring Site Visit

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|--|------|------------------|---------------------|------------------------------------|------------------------------------|
| <u>Very well-aligned implementation^a</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 90.0 | 88.7 | 1.3 * | 0.13 * | 0.047 |
| <i>Corresponding grade equivalent</i> | 6.1 | 5.8 | | | |
| <i>Corresponding percentile</i> | 25 | 22 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 93.2 | 93.2 | 0.0 | 0.00 | 0.984 |
| <i>Corresponding grade equivalent</i> | 7.7 | 7.7 | | | |
| <i>Corresponding percentile</i> | 31 | 31 | | | |
| Sample size | 506 | 370 | | | |
| <u>Well-aligned implementation</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 90.1 | 89.6 | 0.6 | 0.06 | 0.417 |
| <i>Corresponding grade equivalent</i> | 6.1 | 6.0 | | | |
| <i>Corresponding percentile</i> | 25 | 24 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 93.0 | 92.9 | 0.1 | 0.01 | 0.897 |
| <i>Corresponding grade equivalent</i> | 7.7 | 7.7 | | | |
| <i>Corresponding percentile</i> | 31 | 31 | | | |
| Sample size | 430 | 293 | | | |
| <u>Moderately aligned implementation</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 89.9 | 89.7 | 0.2 | 0.02 | 0.837 |
| <i>Corresponding grade equivalent</i> | 6.1 | 6.0 | | | |
| <i>Corresponding percentile</i> | 24 | 24 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 94.1 | 94.4 | -0.2 | -0.03 | 0.757 |
| <i>Corresponding grade equivalent</i> | 7.8 | 7.8 | | | |
| <i>Corresponding percentile</i> | 33 | 34 | | | |
| Sample size | 307 | 236 | | | |

(continued)

Appendix Table I.6 (continued)

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|--|-----|---------------|-----------------------|-----------------------------------|------------------------------|
| <u>Poorly aligned implementation</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | - | - | - | - | - |
| <i>Corresponding grade equivalent</i> | | | | | |
| <i>Corresponding percentile</i> | | | | | |
| Reading vocabulary | | | | | |
| Average standard score | - | - | - | - | - |
| <i>Corresponding grade equivalent</i> | | | | | |
| <i>Corresponding percentile</i> | | | | | |
| Sample size | - | - | | | |
| <hr/> | | | | | |
| | | | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
| Difference in Impacts Between Subgroups | | | | | |
| <u>Very well aligned minus moderately aligned</u> | | | | | |
| Reading comprehension standard score | | | 1.1 | 0.11 | 0.272 |
| Reading vocabulary standard score | | | 0.3 | 0.03 | 0.798 |
| <u>Very well aligned minus well aligned</u> | | | | | |
| Reading comprehension standard score | | | 0.7 | 0.07 | 0.467 |
| Reading vocabulary standard score | | | -0.1 | -0.01 | 0.933 |

(continued)

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

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

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

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

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

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

Appendix Table I.6 (continued)

Rounding may cause slight discrepancies in calculating sums and differences.

^aThe fidelity of program implementation is measured on two dimensions: learning environment and comprehension instruction. On each dimension, schools were measured in terms of their depth of alignment to the program model. Schools that were very well aligned on both dimensions (i.e., with average scores greater or equal to 2.5 on both dimensions) are categorized as having “very well-aligned implementation.” Schools that were well aligned to both dimensions are categorized as being “well-aligned.” Schools that were moderately aligned to at least one dimension and moderately or well aligned to the other dimension are categorized as being “moderately aligned.” Schools that were poorly aligned to one or both dimensions are categorized as being “poorly aligned.”

The group means and impact estimates for the “poorly aligned” category are not reported in the table because only one school was deemed to have poorly aligned implementation fidelity at the second-year spring site visit. In particular, one cannot estimate the impact of the program in “poorly aligned” category schools based on only one school. These results are also suppressed in order to prevent the identification of this particular school and its impact.

The Enhanced Reading Opportunities Study
Appendix Table I.7
Impacts on Reading Behaviors,
Cohort 2 Follow-Up Respondent Sample,
by Program Implementation Fidelity at Spring Site Visit

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value Estimated Impact |
|---|------|------------------|---------------------|------------------------------------|--------------------------------|
| <u>Very well-aligned implementation^a</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 48.0 | 47.2 | 0.8 | 0.02 | 0.761 |
| Amount of non-school-related reading (prior month occurrences) | 31.5 | 31.1 | 0.4 | 0.01 | 0.864 |
| Use of reflective reading strategies (4-point scale) | 2.8 | 2.8 | 0.0 | 0.05 | 0.463 |
| Sample size | 506 | 368 | | | |
| <u>Well-aligned implementation^a</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 43.2 | 41.4 | 1.8 | 0.05 | 0.494 |
| Amount of non-school-related reading (prior month occurrences) | 27.0 | 24.5 | 2.5 | 0.08 | 0.273 |
| Use of reflective reading strategies (4-point scale) | 2.7 | 2.6 | 0.1 | 0.10 | 0.187 |
| Sample size | 429 | 293 | | | |
| <u>Moderately aligned implementation</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 48.4 | 50.5 | -2.0 | -0.05 | 0.526 |
| Amount of non-school-related reading (prior month occurrences) | 32.2 | 34.1 | -1.9 | -0.06 | 0.525 |
| Use of reflective reading strategies (4-point scale) | 2.8 | 2.7 | 0.1 | 0.13 | 0.134 |
| Sample size | 304 | 232 | | | |
| <u>Poorly aligned implementation</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | - | - | - | - | - |
| Amount of non-school-related reading (prior month occurrences) | - | - | - | - | - |
| Use of reflective reading strategies (4-point scale) | - | - | - | - | - |
| Sample size | - | - | | | |

(continued)

Appendix Table I.7 (continued)

| Difference in Impacts Between Subgroups | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
|--|--------------------------|---|---------------------------|
| <u>Very well aligned minus moderately aligned</u> | | | |
| Amount of school-related reading | 2.8 | 0.07 | 0.493 |
| Amount of non-school-related reading | 2.3 | 0.07 | 0.542 |
| Use of reflective reading strategies | -0.1 | -0.08 | 0.427 |
| <u>Very well aligned minus well aligned</u> | | | |
| Amount of school-related reading | -1.1 | -0.03 | 0.776 |
| Amount of non-school-related reading | -2.1 | -0.06 | 0.514 |
| Use of reflective reading strategies | 0.0 | -0.06 | 0.566 |

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

NOTES: The student follow-up survey was administered in spring 2007 at the end of students' ninth-grade year.

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

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 38.322; non-school-related reading standard deviation = 32.976; use of reading strategies standard deviation = 0.592).

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

For each of the above measures, data are missing for no more than 2.2 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

^aThe fidelity of program implementation is measured on two dimensions: learning environment and comprehension instruction. On each dimension, schools were measured in terms of their depth of alignment to the program model. Schools that were very well aligned on both dimensions (i.e., with average scores greater or equal to 2.5 on both dimensions) are categorized as having “very well-aligned implementation.” Schools that were well aligned to both dimensions are categorized as being “well aligned.” Schools that were moderately aligned to at least one dimension and moderately or well aligned to the other dimension are categorized as being “moderately aligned.” Schools that are poorly aligned to one or both dimensions are categorized as being “poorly aligned.”

The group means and impact estimates for the “poorly aligned” category are not reported in the table because only one school was deemed to have poorly aligned implementation fidelity at the second-year spring site visit. In particular, one cannot estimate the impact of the program in “poorly aligned” schools based on only one school. These results are also suppressed in order to prevent the identification of this particular school and its impact.

Appendix Table I.6 shows that the group of schools whose ERO program implementation was very well aligned produced a statistically significant impact on reading comprehension test scores (effect size = 0.13 standard deviation; p-value = 0.047). However, since the difference in impacts across the subgroups of schools defined by implementation fidelity is not statistically significant, it cannot be concluded that the ERO programs were more effective in schools where implementation was rated as very well aligned. Appendix Table I.7 shows that the estimated impact of the programs on reading behaviors is not statistically significant for any of the subgroups of schools defined by implementation fidelity.

Relationship Between Impacts and Number of Weeks to Program Start-Up

Appendix Tables I.8 and I.9 correspond with the third panel of Table 5.6. These tables present impacts on reading test scores and reading behaviors, consecutively, for two groups of schools defined by the number of weeks to program start-up in the second year: those that began operating their ERO program within 2 weeks of the start of the school year and those whose program start-up was delayed by 2 weeks or more.

Appendix Table I.8 shows a statistically significant impact on reading comprehension scores for schools that were able to start operating their program earlier (effect size = 0.10 standard deviation; p-value = 0.048). The difference in impacts across the two subgroups of sites, however, is not statistically significant, and thus it cannot be concluded that the ERO programs were more effective in schools with earlier program start-up. Appendix Table I.9 shows that the impact of the programs on reading behaviors is not statistically significant for either of the two subgroups defined by the number of weeks to program start-up.

To further test the impact of program fidelity and weeks to program start-up on reading comprehension, a composite qualifying statistical test for multiple hypothesis testing was conducted.² This test indicates that the overall variation in impacts across the subgroups defined by implementation fidelity and program start-up is not statistically significant (F-statistic = 0.80; p-value = 0.526), suggesting that the statistical significance of reading comprehension impacts for some of the subgroups in Appendix Tables I.6 through I.9 should be interpreted with caution.³

²This test was conducted by adding a set of interactions between the treatment indicator and the subgroup indicators (three subgroup indicators for implementation fidelity and one indicator for program start-up) to the impact regression model fit to the Cohort 2 follow-up respondent sample. An F-test was then used to test for the joint significance of these interaction terms.

³This overall F-test was also conducted for reading vocabulary, the three reading behaviors outcomes, and the reading behavior composite index. These tests indicate that the overall variation in impacts on these reading outcomes is not statistically significant.

The Enhanced Reading Opportunities Study

Appendix Table I.8

Impacts on Reading Achievement,
Cohort 2 Follow-Up Respondent Sample,
by Number of Weeks Between School Start and ERO Program Start

| Outcome | ERO Group | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|---|-----------|---------------|-----------------------|-----------------------------------|------------------------------|
| <u>Less than 2 weeks</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 90.4 | 89.4 | 1.0 * | 0.10 * | 0.048 |
| <i>Corresponding grade equivalent</i> | 6.2 | 6.0 | | | |
| <i>Corresponding percentile</i> | 25 | 24 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 93.4 | 93.7 | -0.3 | -0.03 | 0.492 |
| <i>Corresponding grade equivalent</i> | 7.7 | 7.8 | | | |
| <i>Corresponding percentile</i> | 32 | 32 | | | |
| Sample size | 881 | 631 | | | |
| <u>2 weeks or more</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 89.5 | 89.1 | 0.4 | 0.04 | 0.546 |
| <i>Corresponding grade equivalent</i> | 6.0 | 5.9 | | | |
| <i>Corresponding percentile</i> | 24 | 23 | | | |
| Reading vocabulary | | | | | |
| Average standard score | 93.7 | 92.9 | 0.7 | 0.08 | 0.324 |
| <i>Corresponding grade equivalent</i> | 7.8 | 7.7 | | | |
| <i>Corresponding percentile</i> | 32 | 31 | | | |
| Sample size | 383 | 276 | | | |
| <u>Difference in impacts</u> | | | | | |
| Difference in Impacts Between Subgroups | | | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
| Reading comprehension standard score | | | 0.5 | 0.05 | 0.534 |
| Reading vocabulary standard score | | | -1.1 | -0.11 | 0.229 |

(continued)

Appendix Table I.8 (continued)

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

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

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

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

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

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

Rounding may cause slight discrepancies in calculating sums and differences.

The Enhanced Reading Opportunities Study
Appendix Table I.9
Impacts on Reading Behaviors,
Cohort 2 Follow-Up Respondent Sample,
by Number of Weeks Between School Start and ERO Program Start

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|---|------|------------------|--------------------------|---|------------------------------------|
| <u>Less than 2 weeks</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 47.2 | 47.1 | 0.1 | 0.00 | 0.952 |
| Amount of non-school-related reading (prior month occurrences) | 30.6 | 30.4 | 0.3 | 0.01 | 0.880 |
| Use of reflective reading strategies (4-point scale) | 2.8 | 2.8 | 0.0 | 0.08 | 0.090 |
| Sample size | 877 | 626 | | | |
| <u>2 weeks or more</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 44.2 | 43.1 | 1.1 | 0.03 | 0.683 |
| Amount of non-school-related reading (prior month occurrences) | 28.2 | 27.1 | 1.0 | 0.03 | 0.679 |
| Use of reflective reading strategies (4-point scale) | 2.7 | 2.6 | 0.1 | 0.10 | 0.217 |
| Sample size | 383 | 275 | | | |
| | | | | | |
| Difference in Impacts Between Subgroups | | | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
| <u>Difference in impacts</u> | | | | | |
| Amount of school-related reading | | | -1.0 | -0.03 | 0.764 |
| Amount of non-school-related reading | | | -0.8 | -0.02 | 0.800 |
| Use of reflective reading strategies | | | 0.0 | -0.02 | 0.873 |

(continued)

Appendix Table I.9 (continued)

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

NOTES: The student follow-up survey was administered in spring 2007 at the end of students' ninth-grade year.

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

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 38.322; non-school-related reading standard deviation = 32.976; use of reflective reading strategies standard deviation = 0.592).

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

For each of the above measures, data are missing for no more than 1.4 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

Relationship Between Impacts and Strong Overall Implementation

Appendix Tables I.10 and I.11 correspond with the final panel in Table 5.6. These tables present the impact estimates by the overall implementation strength of a school's ERO program in the second year of the study. Impacts are presented for the 12 schools whose ERO program was especially strongly implemented as defined by (1) implementation fidelity that was very well aligned to developers' specifications and (2) the program starting up within two weeks of the start of the school year. Impacts are also presented for the 22 schools that did not meet these two conditions.

Table I.10 shows that while ERO programs that were strongly implemented produced a larger impact on reading comprehension than programs that were more weakly implemented (effect size = 0.13 standard deviation and 0.05 standard deviation, respectively), neither of these impacts is statistically significant, nor is the difference between them. Also, while Table I.11 indicates that programs characterized by weaker implementation produced a statistically significant impact on the use of reading strategies (effect size = 0.12 standard deviation; p-value = 0.029), the difference in impacts between the two subgroups of schools is not statistically significant. The composite qualifying test presented in Appendix E indicates that programs that were more weakly implemented did not produce a statistically significant impact on the composite measure of reading behaviors (see Appendix Table E.3). Thus, it cannot be concluded that ERO program effectiveness varied for groups of schools categorized by overall strength of implementation.

The Enhanced Reading Opportunities Study

Appendix Table I.10

Impacts on Reading Achievement,
Cohort 2 Follow-Up Respondent Sample,
by Second-Year Implementation Strength

| Outcome | ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect Size | P-Value for Estimated Impact |
|--|------------|------------------|--------------------------|---|------------------------------------|
| <u>Very well-aligned implementation fidelity and less than 2 weeks to program startup^a</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 90.2 | 88.9 | 1.3 | 0.13 | 0.062 |
| <i>Corresponding grade equivalent</i> | <i>6.1</i> | <i>5.9</i> | | | |
| <i>Corresponding percentile</i> | <i>25</i> | <i>23</i> | | | |
| Reading vocabulary | | | | | |
| Average standard score | 93.2 | 93.4 | -0.2 | -0.02 | 0.729 |
| <i>Corresponding grade equivalent</i> | <i>7.7</i> | <i>7.7</i> | | | |
| <i>Corresponding percentile</i> | <i>31</i> | <i>32</i> | | | |
| Sample size | 464 | 337 | | | |
| <u>Weaker implementation fidelity and/or 2 weeks or more to program startup^b</u> | | | | | |
| Reading comprehension | | | | | |
| Average standard score | 90.1 | 89.5 | 0.5 | 0.05 | 0.296 |
| <i>Corresponding grade equivalent</i> | <i>6.1</i> | <i>6.0</i> | | | |
| <i>Corresponding percentile</i> | <i>25</i> | <i>24</i> | | | |
| Reading vocabulary | | | | | |
| Average standard score | 93.6 | 93.5 | 0.1 | 0.01 | 0.836 |
| <i>Corresponding grade equivalent</i> | <i>7.8</i> | <i>7.8</i> | | | |
| <i>Corresponding percentile</i> | <i>32</i> | <i>32</i> | | | |
| Sample size | 800 | 570 | | | |
| <u>Difference in impacts</u> | | | | | |
| Difference in Impacts Between Subgroups | | | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
| Reading comprehension standard score | | | 0.8 | 0.08 | 0.378 |
| Reading vocabulary standard score | | | -0.3 | -0.03 | 0.689 |

(continued)

Appendix Table I.10 (continued)

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

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

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

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

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

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

Rounding may cause slight discrepancies in calculating sums and differences.

^aThe ERO programs in these schools were deemed to have reached an implementation level that was very well aligned to both the classroom learning environment and comprehension instruction dimensions of the program model by the spring site visit, and their ERO program began operating within 2 weeks of the start of the school year.

^bThe implementation fidelity of the ERO programs in these schools was deemed to be less than very well aligned to the classroom learning environment and/or comprehension instruction dimensions of the program model by the spring site visit, and/or their ERO program began operating 2 weeks or more after the start of the school year.

The Enhanced Reading Opportunities Study

Appendix Table I.11

Impacts on Reading Behaviors,
Cohort 2 Follow-Up Respondent Sample,
by Second-Year Implementation Strength

| Outcome | Non-ERO ERO | Non-ERO Group | Estimated Impact | Estimated Impact Effect | P-Value for Estimated Impact |
|--|----------------|------------------|--------------------------|---|------------------------------------|
| <u>Very well-aligned implementation fidelity and less than 2 weeks to program startup^a</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 48.1 | 46.5 | 1.6 | 0.04 | 0.554 |
| Amount of non-school-related reading (prior month occurrences) | 30.7 | 30.5 | 0.3 | 0.01 | 0.911 |
| Use of reflective reading strategies (4-point scale) | 2.8 | 2.8 | 0.0 | 0.04 | 0.587 |
| Sample size | 464 | 335 | | | |
| <u>Weaker implementation fidelity and/or 2 weeks or more to program startup^b</u> | | | | | |
| Amount of school-related reading (prior month occurrences) | 45.3 | 45.5 | -0.3 | -0.01 | 0.883 |
| Amount of non-school-related reading (prior month occurrences) | 29.4 | 28.8 | 0.6 | 0.02 | 0.751 |
| Use of reflective reading strategies (4-point scale) | 2.7 | 2.7 | 0.1 * | 0.12 * | 0.029 |
| Sample size | 796 | 566 | | | |
| <u>Differences in impact</u> | | | | | |
| Difference in Impacts Between Subgroups | | | Difference in Impacts | Difference in Impact Effect Sizes | P-Value for Difference |
| Amount of school-related reading | | | 1.9 | 0.05 | 0.572 |
| Amount of non-school-related reading | | | -0.3 | -0.01 | 0.919 |
| Use of reflective reading strategies | | | 0.0 | -0.08 | 0.325 |

(continued)

Appendix Table I.11 (continued)

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

NOTES: The student follow-up survey was administered in spring 2007 at the end of students' ninth-grade year.

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

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 38.322; non-school-related reading standard deviation = 32.976; use of reflective reading strategies standard deviation = 0.592).

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

For each of the above measures, data are missing for no more than 2.1 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

^aThe ERO programs in these schools were deemed to have reached an implementation level that was very well aligned to both the classroom learning environment and comprehension instruction dimensions of the program model by the spring site visit, and their ERO program began operating within 2 weeks of the start of the school year.

^bThe implementation fidelity of the ERO programs in these schools was deemed to be less than very well aligned to the classroom learning environment and/or comprehension instruction dimensions of the program model by the spring site visit, and/or their ERO program began operating 2 weeks or more after the start of the school year.

References

- American Guidance Service. 2001a. *Group Reading Assessment and Diagnostic Evaluation: Teacher's Scoring and Interpretive Manual, Level H*. Circle Pines, MN: American Guidance Service.
- American Guidance Service. 2001b. *Group Reading Assessment and Diagnostic Evaluation: Technical Manual*. Circle Pines, MN: American Guidance Service.
- 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, Harold, Carolyn Hill, Alison Rebeck Black, and Mark Lipsey. 2006. "Effect Sizes in Education Research: What They Are, What They Mean, and Why They Are Important." Presentation for the Institute of Education Sciences 2006 Research Conference. New York: MDRC.
- Cronbach, Lee J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16: 297-334.
- Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. *Using Randomization in Development Economics: A Toolkit*. London: Centre for Economic Policy Research.
Web site: <http://www.cepr.org/pubs/dps/DP6059.asp>.
- Fine, Michelle. 1988. "Deinstitutionalizing Educational Inequity." Pages 88-119 in Council of Chief State School Officers (eds.), *School Success for Students at Risk*. New York: Harcourt.
- Hedges, Larry V., and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis*. San Diego, CA: Academic Press.
- Diana Oxley (ed.). *Critical Issues in Development and Implementation of High School Small Learning Communities*. Washington, DC: U.S. Department of Education.
- Kempe, 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., and Jeffrey B. Liebman, 2004. "Experimental Analysis of Neighborhood Effects on Youth." KSG Working Paper RWP04-034. Cambridge, MA: Harvard University, John F. Kennedy School of Government.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75, 1: 83-119.
- Lipsey, Mark. 1990. *Design Sensitivity: Statistical Power for Experimental Research*. Newbury Park, CA: Sage.
- Logan, Brent R., and Ajit C. Tamhane. 2003. "On O'Brien's OLS and GLS Tests for Multiple Endpoints." Department of Industrial Engineering and Management Sciences Working Paper 03-0004. Evanston, IL: Northwestern University.
- 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>.
- O'Brien, Peter C. 1984. "Procedures for Comparing Samples with Multiple Endpoints." *Biometrics* 40: 1079-1087.
- Roderick, Melissa. 1993. *The Path to Dropping Out: Evidence for Intervention*. Westport, CT: Auburn House.
- Schoenbach, Ruth, Cynthia Greenleaf, Christine Cziko, and Lori Hurwitz. 1999. *Reading for Understanding: A Guide to Improving Reading in Middle and High School Classrooms*. San Francisco: Jossey-Bass.
- Schumaker, Jean B., and Donald D. Deshler. 2003. "Designs for Applied Educational Research." Pages 283-500 in H. Lee Swanson, Karen R. Harris, and Steve Graham (eds.), *Handbook of Learning Disabilities*. New York: Guilford.
- Schumaker, Jean B., and Donald D. Deshler. 2004. "Teaching Adolescents to be Strategic Learners." In Donald D. Deshler and Jean B. Schumaker (eds.), *High School Students with Disabilities: Strategies for Accessing the Curriculum*. New York: Corwin.
- U.S. Department of Education. 2005. *Smaller Learning Communities: Special Competition for Supplemental Reading Program Research Evaluation*. Washington DC: U.S. Department of Education.
- WestEd. 2004. "High Quality Supplemental Literacy Programs for Striving Readers in Ninth Grade." Unpublished proposal. San Francisco: WestEd.
- Wigfield, Allen. 2004. "Motivation for Reading During the Early Adolescent and Adolescent Years." Pages 56-69 in Dorothy Strickland and Donna Alvermann (eds.), *Bridging the Literacy Achievement Gap, Grades 4-12*. New York: Teachers College Press.