# Portland Public Schools' Striving Readers Program

*Year 5 Evaluation Report*

Prepared for

**U.S. Department of Education**
Office of Elementary and Secondary Education
400 Maryland Avenue, Room 3C118
Washington, DC 20202

Prepared by

**RMC Research Corporation**
111 SW Columbia Street, Suite 1200
Portland, OR 97201

**December 22, 2011**

# Portland Public Schools' Striving Readers Program

*Year 4 Evaluation Report*

*Prepared for*

**U.S. Department of Education**
Office of Elementary and Secondary Education
400 Maryland Avenue, Room 3C118
Washington, DC 20202

*Prepared by*

**Bonnie J. Faddis, Project Director**
**Margaret Beam**
**Lauren Maxim**
**Elizabeth Vale Gandhi**
**Karen Hahn**
**Rita Hale**

**RMC Research Corporation**
111 SW Columbia Street, Suite 1200
Portland, OR 97201

*December 22, 2011*

RMC RESEARCH

# Contents

# Exhibits

# Executive Summary

Portland Public Schools (PPS), the largest school district in Oregon, serves more than 46,000 students in regular and special programs. More than 2,900 classroom teachers address the needs of a diverse student population (44% minority, 46% low income, 14% special education, 9% English language learners). A district needs assessment in fall 2005 revealed that 13 of Portland's 85 regular schools were eligible to participate in the Striving Readers program. Four of the high schools and 5 of the middle schools determined that they could meet the program's research requirements. All 9 schools had at each grade level a significant number of students who were at least 2 years behind in reading achievement; all received Title I funding; and none had achieved Adequate Yearly Progress under No Child Left Behind at the time of the Striving Readers application. School leaders expected the Striving Readers program to impact more than 6,400 students and 450 teachers in the 9 participating schools.

After examining adolescent reading programs and studying the research on adolescent literacy, Portland Public Schools selected the Strategic Instruction Model Content Literacy Continuum developed by the University of Kansas' Center for Research on Learning to improve teacher instruction and student reading achievement in the participating middle and high schools. This report summarizes Year 1 (2006–2007), Year 2 (2007–2008), Year 3 (2008–2009), Year 4 (2009–2010), and Year 5 (2010–2011) of implementation of the targeted intervention for students reading at least 2 years below grade level in Grades 7–10 and the whole school intervention designed to help all students in Grades 6–12 learn the critical content in all curricular areas.

## Targeted Intervention

The Xtreme Reading curriculum serves as the targeted intervention for Portland Public Schools' Striving Readers program. Xtreme Reading is designed for middle and high school students who need explicit strategy instruction to develop the reading skills needed to master critical course content. Xtreme Reading focuses on 6 reading strategies: Word Mapping, Word Identification, Self-Questioning, Visual Imagery, Paraphrasing, and Inference.

To evaluate the implementation of the targeted intervention, the evaluation team collected data from the professional developers, the Striving Readers district leadership team, school

administrators, school literacy coaches, Xtreme Reading teachers, and classroom observations. The data were analyzed in terms of professional development inputs, classroom implementation fidelity, and teacher buy-in. Although both school levels showed improvement from Year 1 to Year 5, participation in the professional development offered and teacher qualifications were generally higher in the middle schools than the high schools in all years. For example, in Year 5, 85% of the middle school teachers and 75% of the high school teachers attended at least three fourths of the professional development sessions offered. Classroom implementation fidelity and teacher buy-in were also higher in the middle schools than the high schools in all years. One third of the middle schools but none of the high schools were rated as achieving high classroom implementation fidelity in Year 5; one middle school and 2 high schools were judged to have low classroom implementation fidelity. One half of the middle schools and one quarter of the high schools were rated as having high levels of teacher buy-in. All Xtreme Reading classrooms were staffed by a pair of teachers in Year 1; in Year 2 this proportion decreased to 67%; and in Years 3 through 5 all Xtreme Reading classes were taught by a single teacher.

### *Impact of Xtreme Reading*

The impact evaluation of Xtreme Reading was designed to answer 5 questions:

- To what extent do students in Xtreme Reading improve their reading skills after one year compared to the control group?
- Do the effects of Xtreme Reading differ by school level?
- Do the effects of Xtreme Reading differ by cohort?
- What percentage of students in Xtreme Reading are reading at grade level by the end of the school year?
- To what extent do students in Xtreme Reading improve their reading motivation after one year compared to the control group?

The impact evaluation examined differences between the treatment and control group students in terms of reading achievement on the Group Reading Assessment and Diagnostic Evaluation (GRADE), which was administered each year of Years 1 through 4 in the fall and spring. The combined analytic sample for Years 1 through 4 included 756 treatment group and 823 control group students who had posttest scores. At each time point students in both groups were also asked to complete a survey that measured their motivation for reading. Student outcome data

for Year 5 are not included in this analysis because the district elected to discontinue the experimental design for the final year of funding. Instead, the district funded 9 new K-8 schools and 1 new middle school to implement the Xtreme Reading program, and all 20 Striving Readers schools were permitted to serve any student who qualified for the program.

A multilevel model was used to estimate the impact of the targeted intervention on spring GRADE normal curve equivalent (NCE) scores while controlling for fall GRADE NCE scores, grade level, ethnicity, and English language proficiency. The analysis for Years 1 through 4 revealed a significant intervention effect for the overall sample: the treatment group students had significantly higher outcomes than the control group students. A significant treatment effect was also present for both the middle school and high school samples. The standardized effect size for the overall sample was .21; at the middle school level the standardized effect size was .29, and at the high school level the standardized effect size was .12.

Approximately 60% of both the treatment and control group students were reading 2 or more years below grade level at the time of the GRADE pretest. Of this group, only 4% of the 454 treatment group students and 2% of the 492 control group students were reading at or above grade level at posttest. Of all 756 treatment and 823 control group students, only 92 treatment and 56 control group students (12% and 7%, respectively) were reading at or above grade level after the intervention year. The same multilevel model used in the main impact analysis was used to estimate the impact of the intervention on spring Oregon Assessment of Knowledge and Skills (OAKS) NCEs. The data analysis revealed no intervention effect for the high school sample; however significant treatment effects were present for the overall sample ($\beta = 1.26$, $p < .05$) and the middle school sample ($\beta = 1.69$, $p < .05$).

No differences between the treatment and control groups were evident in terms of change in reading motivation for Cohort 1. However, in Cohorts 2–4 combined the treatment group students gained significantly more than the control group students in overall reading motivation. When the results were examined separately for middle and high school students, the differences between the treatment and control groups were not significant. The differences in program impact by school level are not surprising given the differences in program implementation: the middle schools had higher fidelity of Xtreme Reading implementation in the classroom and higher levels of teacher buy-in.

## Whole School Intervention

Content enhancement routines, which help students understand the key content in all of their courses, serve as the whole school intervention for Portland Public Schools' Striving Readers program. In Year 1 language arts and social studies teachers received training on the use of the content enhancement routines: Unit Organizer, Framing, Vocabulary LINCing, and Concept Mastery. In Year 2 math teachers were introduced to the Unit Organizer and Framing content enhancement routines, and science teachers were introduced to the Framing, Concept Mastery, and the new Chapter Survey routines. In Year 3 arts, physical education, and health teachers received training on the use of the Unit Organizer, Framing, and Order routines; in addition, the arts teachers were trained on the Concept Mastery routine and the physical education and health teachers were trained on the Concept Comparison routine. The professional developers also offered optional training on the Question Exploration, Clarifying, Order, Lesson Organizer, and Course Organizer content enhancement routines. In Years 4 and 5, new and continuing content teachers had the option of participating in summer training to either learn the basic content enhancement routines or learn to teach embedded reading strategies and advanced content enhancement routines.

To provide context for interpreting the impact of the whole school intervention on student reading achievement, the evaluation team assessed the implementation of the content enhancement routines each year during Years 1 through 4. The data were analyzed in terms of professional development inputs, classroom implementation fidelity, and teacher buy-in. Overall, the middle school teachers participated in the professional development more than the high school teachers. The middle school teachers also viewed the content enhancement routines more favorably than did the high school teachers and were more likely to allow classroom observations and to work with the school literacy coaches. In Year 4 classroom implementation fidelity was rated medium at most of the middle schools and low at most of the high schools.

In Year 1 school administrators believed that the primary barriers to implementation included their own lack of knowledge of the whole school intervention features, the competing demands of multiple reforms and programs, the late hiring of key staff, some teachers' unwillingness to participate in the Striving Readers program, and conflicts between the professional development requirements for Striving Readers and other initiatives. Principals believed that implementation would have been more successful if all teachers (rather than language arts and social studies teachers only) had been involved in the program in Year 1 and issues involving the teachers

union had been resolved earlier. In Year 2 the district leadership team, school administrators, and the professional developers all noted improved teacher attitudes toward Striving Readers, although many of the same implementation barriers were in effect. In Year 3 teacher attitudes toward the Striving Readers program continued to improve but issues persisted with respect to the quality and consistency of the professional development provided by the external professional developers. By Year 4, math, science and PE teachers were still asking for more specific examples of content enhancement routines for their content areas; teachers wanted more time to see content enhancement routines their colleagues had constructed and more paid time to develop content enhancement routines; and teachers continued to complain about the GIST software for creating content enhancement routines on their computers.

### *Impact of the Content Enhancement Routines*

The whole school intervention impact evaluation addressed the extent to which students in Grades 6–10 improved their OAKS reading scores after implementation of the content enhancement routines. Average school OAKS scores were tracked over time beginning 3 years prior to implementation of the Striving Readers grant through spring 2010. The final year of implementing the whole school intervention was not included in this analysis because of changes in the OAKS test and discontinuation of the whole school intervention in the high schools. The data analysis revealed no significant gains in OAKS total reading scores from pre- to postimplementation. This finding was not unexpected due to the overall weak implementation of the intervention, multiple school reconfigurations and staffing changes, and changes in district curricula and OAKS content over the pre- and postimplementation period.

# Introduction

Portland Public Schools (PPS), the largest school district in Oregon, serves more than 46,000 students in regular and special programs. More than 2,900 classroom teachers address the needs of a diverse student population (44% minority; 45% low income; 14% special education; 10% English language learners). A district needs assessment conducted in fall 2005 revealed that 13 of Portland's 85 regular schools were eligible to participate in the Striving Readers program: 5 high schools and 8 middle schools. All of these schools received Title I funding, and none had achieved Adequate Yearly Progress under No Child Left Behind at the time of the Striving Readers application in 2005.

Four of the high schools and 5 of the middle schools determined that they could meet the research requirements of the Striving Readers program. All 9 schools had a significant number of students at each grade level who were at least 2 years behind in reading achievement. Portland Public Schools was just beginning to recover from the neglect to its curriculum and professional development resources caused by more than 10 years of funding reductions and drastic staff cutbacks. Schools were, however, committed to improving reading instruction for struggling readers and embedding reading strategies into content instruction. School leaders expected the Striving Readers program to impact more than 6,400 students and 450 teachers in the 9 participating schools.

After examining adolescent reading programs and studying the research on adolescent literacy, Portland Public Schools selected the Strategic Instruction Model Content Literacy Continuum developed by the University of Kansas' Center for Research on Learning to improve teacher instruction and student reading achievement in the participating middle and high schools. This report summarizes Year 1 (2006–2007), Year 2 (2007–2008), Year 3 (2008–2009), and Year 4 (2009–2010) of the implementation of the targeted intervention for students reading at least 2 years below grade level in Grades 7–10. The report also describes the whole school intervention designed to help all students in Grades 6–12 learn the critical content in all curricular areas.

## Intervention Models

The Content Literacy Continuum developed by the University of Kansas' Center for Research on Learning comprises both a targeted intervention and a whole school intervention.

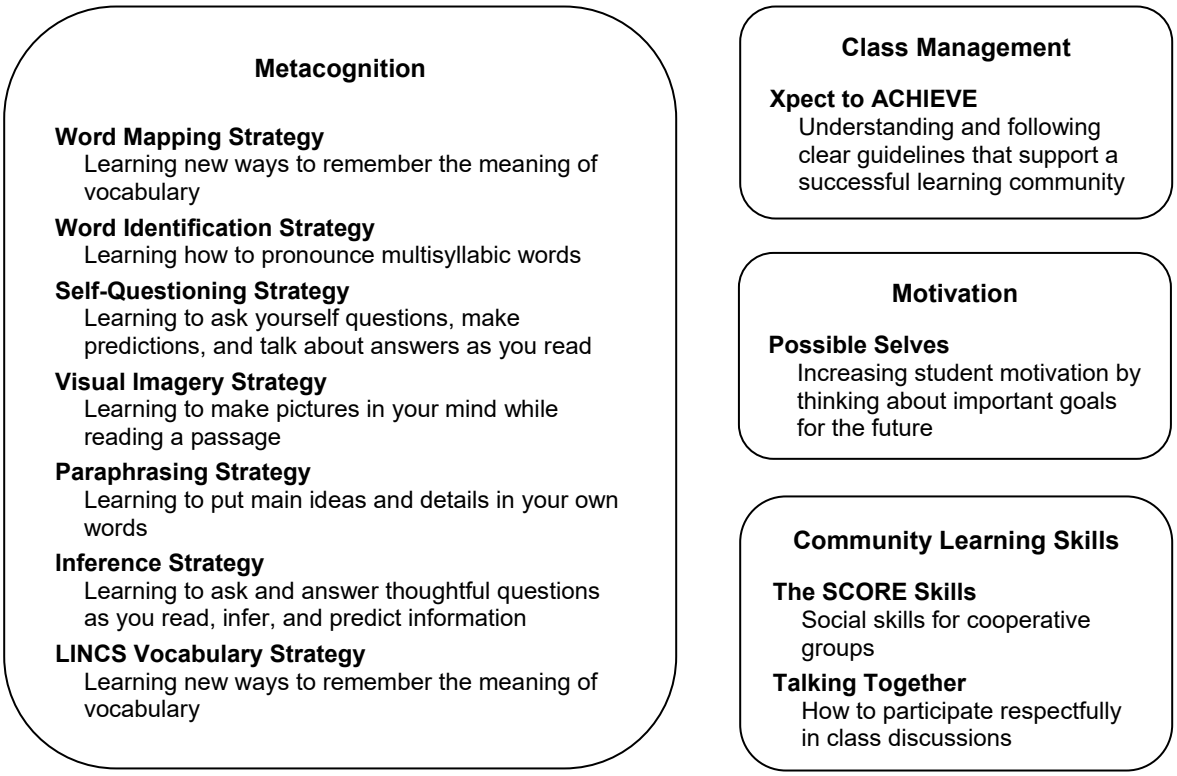
### *Targeted Intervention*

Portland Public Schools' Striving Readers program utilizes the Xtreme Reading curriculum for the targeted intervention. Xtreme Reading is a 1-year program designed for middle and high school students who need explicit strategy instruction to develop the reading skills needed to master critical course content. In Year 4 the Xtreme Reading program focused on 6 reading strategies: Word Mapping, Word Identification, Self-Questioning, Visual Imagery, Paraphrasing, and Inference. The LINCs Vocabulary strategy implemented in Years 1 and 2 of the Striving Readers grant was eliminated to allow more time for teaching the other 6 strategies. Xtreme Reading's core instructional approaches include direct instruction, teacher modeling, paired student practice, and independent practice. Reading motivation and self-directed learning are encouraged through collaborative learning, self-selection of highly engaging texts, and teacher think-aloud modeling. Exhibit 1 summarizes the components and strategies of Xtreme Reading. The curriculum includes trade books that are used to practice the reading strategies and encourage independent reading and a student binder that incorporates strategy cue cards, worksheets, and fluency/comprehension tests.

### *Whole School Intervention*

For the whole school intervention Portland Public Schools' Striving Readers program utilizes content enhancement routines, which help students understand the key content in all of their courses. In Year 1 language arts and social studies teachers received training on the use of content enhancement routines. Math and science teachers received training in Year 2, and arts, physical education, and health teachers received training in Year 3. Special education, ESL, and foreign language teachers participated at each school's discretion. Exhibit 2 summarizes the content enhancement routines that are part of the Content Literacy Continuum. The content enhancement routines are designed to help academically diverse classes focus on key content and involve students in the instructional process.

## Exhibit 1
## Xtreme Reading Model




*Note.* Information provided by the Center for Research on Learning at the University of Kansas, 2007.

## Exhibit 2
## Content Enhancement Routines

| **Planning and Leading Learning** | **Teaching Concepts** |
|---|---|
| ▪ Course Organizer | ▪ Concept Mastery |
| ▪ Unit Organizer | ▪ Concept Anchoring |
| ▪ Lesson Organizer | ▪ Concept Comparison |
| **Exploring Text, Topics, and Details** | **Increasing Performance** |
| ▪ Framing | ▪ Quality Assignment |
| ▪ Survey | ▪ Question Exploration |
| ▪ Clarifying | ▪ Recall Enhancement |
| ▪ Order | ▪ Vocabulary LINCing |

*Note.* Data provided by the Center for Research on Learning at the University of Kansas, November 2007.

All of the content enhancement routines follow a 3-step sequence of Cue, Do, and Review. This sequence draws students' attention to the instructional process specific to each routine, involves students in the process, and assesses student understanding of the process. In Year 1 all teachers were introduced to 4 content enhancement routines:

- **Unit Organizer**—Introduces a content unit and shows how concepts are related. The Unit Organizer routine is typically constructed by the teacher and students at the beginning of the unit.
- **Framing**—Helps students conceptualize the main ideas of a topic. The teacher and students construct the Framing routine by completing sections for the key topic, the "is about . . ." section, the main ideas, the essential details, and a "so what" statement.
- **Vocabulary LINCing**—Uses auditory and visual memory devices to help students learn new vocabulary words. The teacher and students construct a LINCs table that includes 5 key components: the term, the reminding word, the LINCing story, the LINCing picture, and the definition.
- **Concept Mastery**—Helps students master key concepts and see how the concepts relate to the larger body of knowledge. The teacher and students construct a concept diagram that displays a definition, key points, and good and bad examples.

In Year 2 math teachers were introduced to the Unit Organizer and Framing content enhancement routines, and science teachers were introduced to the Framing, Concept Mastery, and Chapter Survey routines. The Chapter Survey routine was new:

- **Chapter Survey**—Helps students analyze a passage of text and identify key information. The teacher and students complete a worksheet that results in a graphic depiction of the structure of the reading passage.

The professional developers also offered optional training on the following content enhancement routines:

- **Concept Anchoring**—Associates new concepts to familiar concepts to aid understanding. The teacher and students construct an anchoring table that explores the distinct and shared characteristics of the new and the known concepts.
- **Concept Comparison**—Helps students understand related concepts through comparison of their similarities and differences. The teacher and students conduct the comparison though the construction of the concept comparison table.

- **Lesson Organizer**—Contextualizes the unit within the course. The teacher and students construct the lesson organizer at the beginning of the lesson.
- **Course Organizer**—Helps students understand the critical content that needs to be learned in the course, how units within the course have been organized, and the ongoing routines and strategies that will be used.

In Year 3 arts, physical education, and health teachers received training on the use of the Unit Organizer, Framing, and Order routines; in addition, the arts teachers were trained on the Concept Mastery routine and the physical education and health teachers were trained on the Concept Comparison routine. Over the course of the year the professional developers at each school offered facilitated work time to help interested teachers practice integrating content enhancement routines in their lesson planning. The professional developers also offered optional training on the Lesson Organizer and Course Organizer routines as well as the following content enhancement routines:

- **Question Exploration**—Helps students show what they know by exploring a critical question, key terms related to the critical question, supporting questions and answers, and a main idea that synthesizes the answers to these questions.
- **Clarifying**—Helps students master the meaning of new words and phrases by connecting them to their own knowledge and experiences and using the new words in a variety of ways.
- **Order**—Helps students decide what information is important and express it in a format that is useful for the student.

Beginning in Year 3, the professional developers also offered training on several embedded reading strategies: Word Mapping, Word Identification, Self-Questioning, Visual Imagery, and Paraphrasing. The professional developers also offered a series of training institutes for teachers interested in becoming certified Strategic Instruction Model trainers. In Year 4, most of the professional development was provided by the PPS certified trainers and included sessions for teachers who were new to Content Enhancement Routines as well as teachers with previous experience. The professional development for experienced teachers focused on embedded strategies, integrated units, and content enhancement routines that had not been previously taught at all schools. These content enhancement routines included the Course Organizer, Clarifying, Chapter Survey, and Question Exploration routines as well as 2 new content enhancement routines:

- **Quality Assignment—**Helps students improve the quality of their assignments and the rate of completion by ensuring that the assignment has a clear purpose, is authentic, is personally relevant, is optimally challenging, represents variety, promotes creativity and interaction, allows students to make choices, and has clear directions and grading criteria.
- **Recall Enhancement—**Helps students recall new information using different types of devices such as Snapshots, Boxing, and Acronyms.

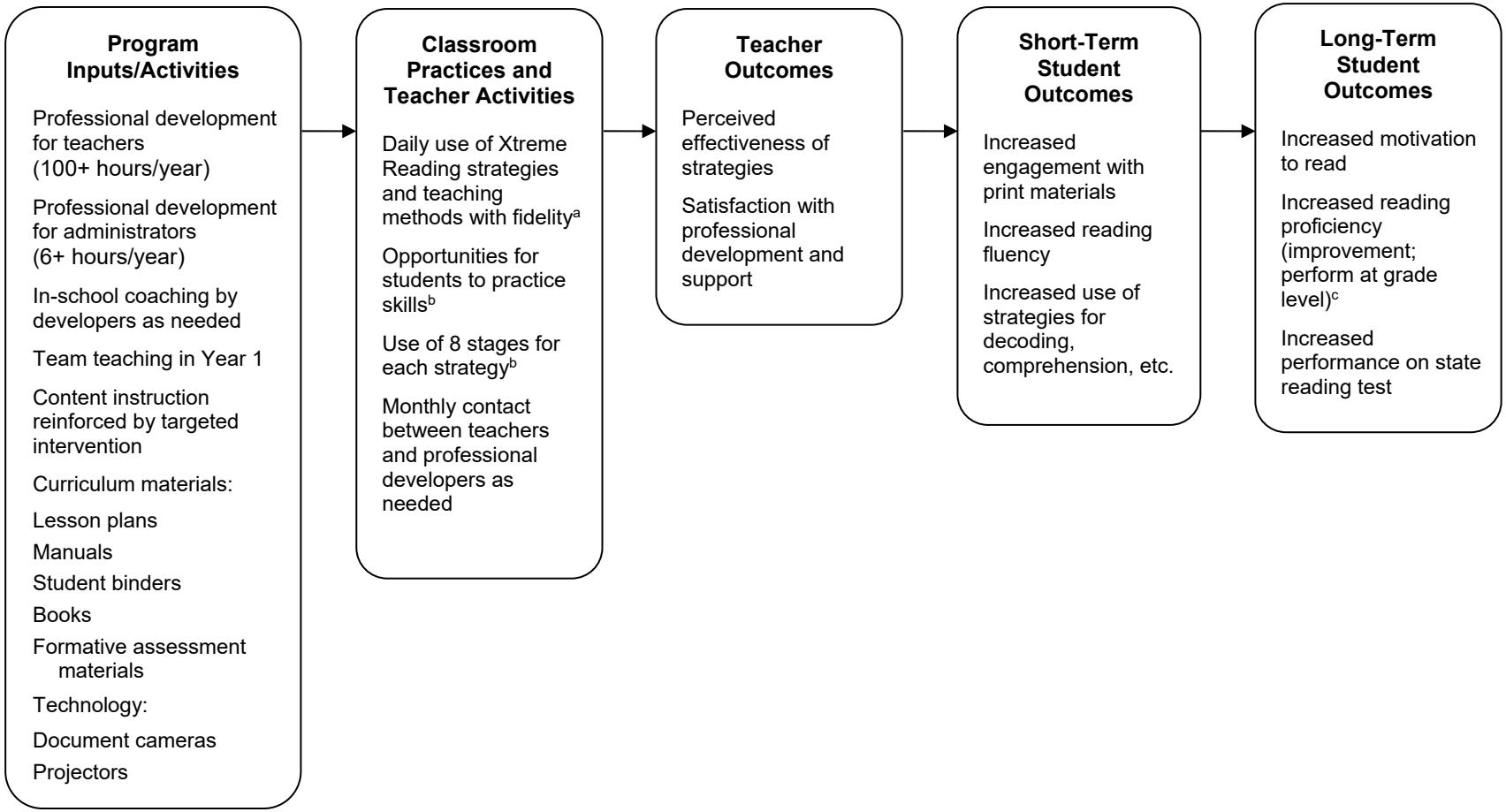
The professional development for new teachers covered the more basic content enhancement routines such as unit organizer, frame, and concept mastery routines.

In Year 5, district Striving Readers staff provided facilitated work time at the 6 middle schools and 1 of the high schools to assist content teachers in developing lessons using content enhancement routines. District staff also provided training on specific routines as requested.

## Logic Models

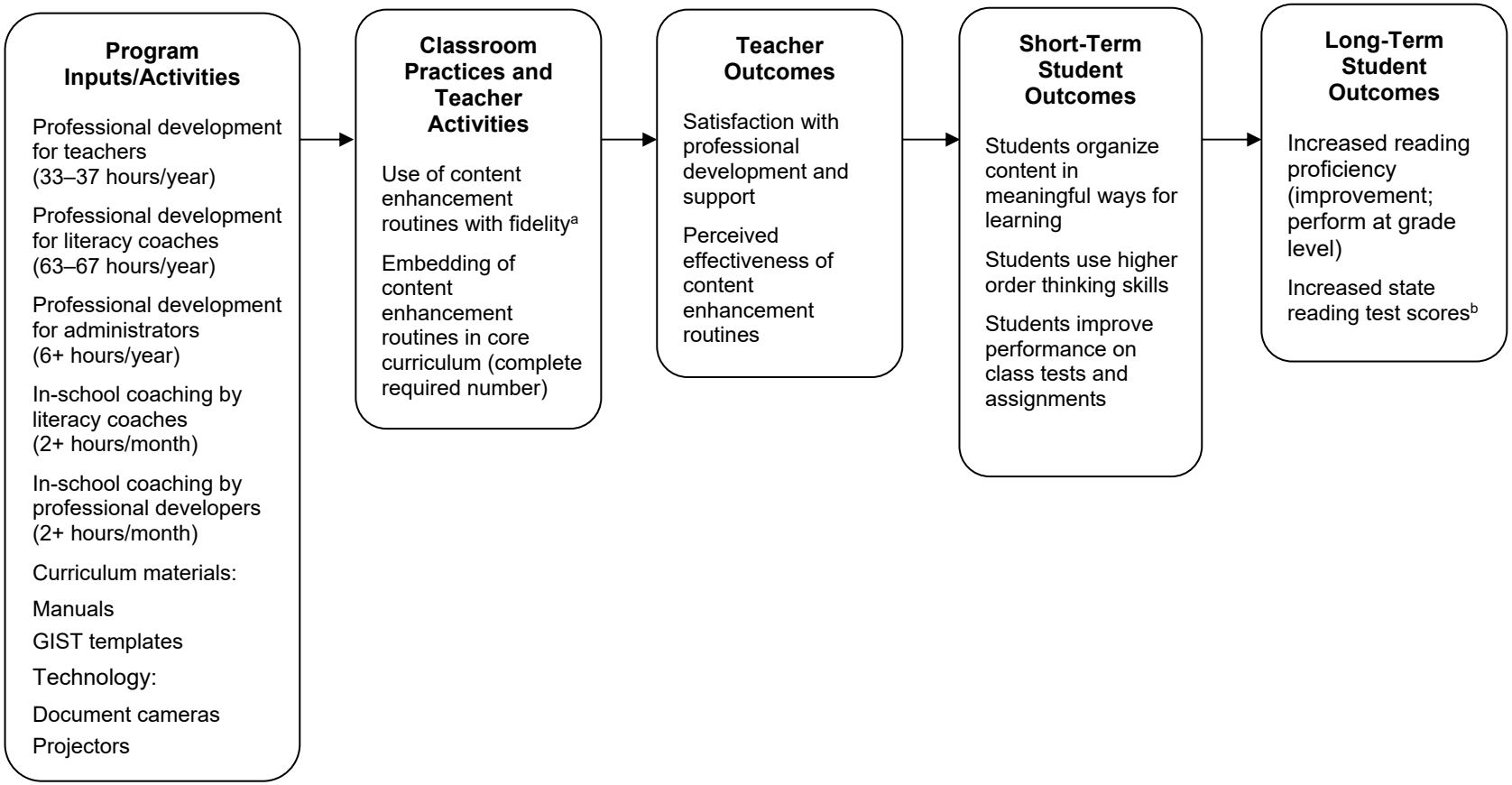
Exhibit 3 presents the logic model for the targeted intervention, and Exhibit 4 presents the logic model for the whole school intervention.

**Exhibit 3**
**Logic Model for Targeted Intervention**




[a]See Exhibit 1 for description of Xtreme Reading curriculum. [b]See Appendix A for stages of Xtreme Reading strategy instruction. [c]District goals at beginning of project were for 80% of students in the targeted intervention to read at or above their grade level after one year of program participation.

**Exhibit 4**
**Logic Model for Whole School Intervention**




[a]See Exhibit 2 for description of content enhancement routines. [b]District goals at beginning of project were that the percentage of students at benchmark level on the OAKS would increase by 10–15% annually.

## Overview of the Evaluation Design

Exhibit 5 presents research questions and data sources linked to the research questions for the targeted intervention. Exhibit 6 presents research questions and data sources for the whole school intervention.

**Exhibit 5**
**Data Sources for the Targeted Intervention**

| Research Question | Surveys and Interviews: Developer | Surveys and Interviews: District Staff | Surveys and Interviews: Principal | Surveys and Interviews: Coach | Surveys and Interviews: Teacher | Classroom Observation | PPS Records Review |
|---|---|---|---|---|---|---|---|
| **Professional Development for Teachers (Xtreme Reading)** | | | | | | | |
| 1. **To what extent did teachers participate in group professional development?** | | | | | | | |
| **a. How many hours of professional development were offered?** | | ◆ | | | | | ◆ |
| **b. What topics were offered?** | ◆ | ◆ | | | ◆ | | ◆ |
| **c. What percent of Tier I teachers participated in each phase of professional development?** | ◆ | ◆ | | | | | ◆ |
| **d. To what extent were teachers satisfied with the group professional development they received?** | | ◆ | ◆ | ◆ | ◆ | | |
| 2. **What do teachers believe are the primary outcomes from the professional development they received?** | | | | | | | |
| **a. How has their teaching changed?** | | | | ◆ | ◆ | | |
| **b. How has student learning changed?** | | | ◆ | ◆ | ◆ | | |
| 3. **How could professional development for teachers be improved?** | | ◆ | ◆ | ◆ | ◆ | | |
| **Professional Development for School Literacy Coaches** | | | | | | | |
| 4. **To what extent did school literacy coaches participate in group professional development?** | | | | | | | |
| **a. How many hours of professional development were offered?** | | ◆ | | | | | ◆ |
| **b. What topics were offered?** | ◆ | ◆ | | | | | ◆ |
| **c. What percent of literacy coaches participated in each phase of professional development?** | | ◆ | | | | | ◆ |
| **d. To what extent were school literacy coaches satisfied with the group professional development they received?** | | ◆ | ◆ | ◆ | | | |

*(exhibit continues)*

**Exhibit 5 (continued)**

| Research Question | Surveys and Interviews: Developer | District Staff | Principal | Coach | Teacher | Classroom Observation | PPS Records Review |
|---|---|---|---|---|---|---|---|
| 5. **How could professional development for school literacy coaches be improved?** | ◆ | ◆ | ◆ | ◆ | | | |
| 6. **To what extent did school literacy coaches participate in individual professional development with either University of Kansas or district staff?** | ◆ | | ◆ | ◆ | | | |
| Professional Development for School Administrators | | | | | | | |
| 7. **To what extent did school administrators participate in group professional development?** | | | | | | | |
| a. **How many hours of professional development were offered?** | | ◆ | | | | | ◆ |
| b. **What topics were offered?** | | ◆ | | | | | ◆ |
| c. **What percent of school administrators participated in each phase of professional development?** | | ◆ | | | | | ◆ |
| d. **To what extent were school administrators satisfied with the group professional development they received?** | | ◆ | ◆ | | | | |
| 8. **How could professional development for school administrators be improved?** | | ◆ | ◆ | ◆ | | | |
| Professional Development Summary | | | | | | | |
| 9. **What factors contributed to the successful implementation of professional development?** | | ◆ | | | | | |
| Instructional Coaching | | | | | | | |
| 10. **To what extent did teachers participate in coaching?** | | | | | | | |
| a. **What percent of teachers worked with their school literacy coach?** | | | | ◆ | ◆ | | |
| b. **What activities did teachers work on with their school literacy coach?** | | | | ◆ | ◆ | | |
| c. **How much time did teachers spend, on average, with their school literacy coach?** | | | | ◆ | ◆ | | |
| d. **To what extent were teachers satisfied with the coaching they received from their school literacy coach?** | | | | ◆ | ◆ | | |
| e. **What percent of teachers worked with the University of Kansas professional developers?** | ◆ | | | | ◆ | | |
| f. **What topics did teachers work on with the University of Kansas professional developers?** | ◆ | | | | ◆ | | |

*(exhibit continues)*

**Exhibit 5 (continued)**

| Research Question | Surveys and Interviews: Developer | District Staff | Principal | Coach | Teacher | Classroom Observation | PPS Records Review |
|---|---|---|---|---|---|---|---|
| g. **How much time did teachers spend, on average, with the University of Kansas professional developers?** | ◆ | | | | ◆ | | |
| h. **To what extent were teachers satisfied with the coaching they received from the University of Kansas professional developers?** | ◆ | | | ◆ | ◆ | | |
| 11. **To what extent did school literacy coaches implement the specified instructional coaching activities?** | ◆ | | | ◆ | | | |
| 12. **How does coaching vary with teacher content area?** | | | | ◆ | | | |
| 13. **How does coaching vary for content and Xtreme teachers?** | | | | ◆ | | | |
| District- and School-Level Administrative Support | | | | | | | |
| 14. **What types of support has the district office provided to Striving Readers schools?** | | ◆ | ◆ | ◆ | ◆ | | |
| 15. **What additional support is needed from the district office?** | | ◆ | ◆ | ◆ | ◆ | | |
| 16. **What types of support have administrators provided to Striving Readers teachers and coaches?** | | | ◆ | ◆ | ◆ | | |
| 17. **What additional support is needed from school administrators?** | | ◆ | | ◆ | ◆ | | |
| Staff Characteristics | | | | | | | |
| 18. **What recruitment strategies were used to find teachers and coaches for Striving Readers?** | | ◆ | ◆ | | | | |
| 19. **What training and experience do teachers and coaches have?** | | | | | ◆ | | |
| 20. **To what extent did teacher and coach characteristics meet desired criteria?** | | ◆ | ◆ | | | | |
| Classroom Materials and Equipment | | | | | | | |
| 21. **What materials, software, and equipment were Xtreme teachers expected to have?** | | ◆ | | | ◆ | | |
| 22. **To what extent did teachers have the materials, software, and equipment needed for successful implementation?** | | ◆ | | | ◆ | | |
| 23. **In what ways do Xtreme teacher materials, software, and equipment need to be improved?** | | ◆ | | ◆ | ◆ | | |
| 24. **How useful are the materials, software, and equipment for implementing the program?** | | ◆ | | ◆ | ◆ | | |

*(exhibit continues)*

**Exhibit 5 (continued)**

| Research Question | Surveys and Interviews: Developer | District Staff | Principal | Coach | Teacher | Classroom Observation | PPS Records Review |
|---|---|---|---|---|---|---|---|
| **Instructional Processes** | | | | | | | |
| 25. **To what extent did the Xtreme teachers implement the program with fidelity?** | | | | | | | |
| a. **Which components of the program were implemented with fidelity?** | ◆ | | | | | ◆ | |
| b. **Which components of the program were not implemented with fidelity?** | ◆ | | | | | ◆ | |
| c. **What were the barriers to successful implementation of Xtreme Reading?** | ◆ | ◆ | ◆ | ◆ | ◆ | | |
| d. **What steps are University of Kansas and district staff taking to improve fidelity of program implementation?** | ◆ | ◆ | | | | | |
| e. **What is the nature of the coteaching relationship?** | | | | | | ◆ | |
| **Formative Assessment** | | | | | | | |
| 26. **What assessment activities are integral to the Xtreme reading program?** | | ◆ | | | | | |
| a. **To what extent were these assessment activities implemented?** | ◆ | ◆ | | | ◆ | | |
| b. **In what ways are teachers expected to use assessment data?** | ◆ | ◆ | | | ◆ | | |
| **Implementation Summary** | | | | | | | |
| 27. **Overall, in what important ways did the program as implemented differ from the program as planned?** | | ◆ | | | | | |
| 28. **To what extent did fidelity of implementation differ across schools, teachers, and years?** | ◆ | ◆ | ◆ | | ◆ | | |
| 29. **What is the overall level of satisfaction with the Striving Readers program?** | ◆ | ◆ | ◆ | ◆ | ◆ | | |
| **Impacts of Targeted Intervention** | | | | | | | |
| 30. **To what extent do students in Xtreme Reading improve their reading skills after one year compared to the control group?** | | | | | | | ◆ |
| 31. **Do the effects of Xtreme Reading differ for middle and high school students?** | | | | | | | ◆ |
| 32. **Do the effects of Xtreme Reading differ for each cohort?** | | | | | | | ◆ |
| 33. **What percentage of students in Xtreme Reading reach grade level by the end of the school year?** | | | | | | | ◆ |
| 34. **To what extent do students in Xtreme Reading improve their reading motivation after one year compared to the control group?** | | | | | | | ◆ |

**Exhibit 6**
**Data Sources for the Whole School Intervention**

| Research Question | Surveys and Interviews: Developer | District Staff | Principal | Coach | Teacher | Classroom Observation | PPS Records Review |
|---|---|---|---|---|---|---|---|
| **Professional Development for Teachers (Content)** | | | | | | | |
| 1. **To what extent did teachers participate in group professional development?** | | | | | | | |
| **a. How many hours of professional development were offered?** | | ◆ | | | | | ◆ |
| **b. What topics were offered?** | ◆ | ◆ | | | ◆ | | ◆ |
| **c. What percent of content teachers participated in each phase of professional development?** | ◆ | ◆ | | | | | ◆ |
| **d. To what extent were teachers satisfied with the group professional development they received?** | | ◆ | ◆ | ◆ | ◆ | | |
| 2. **What do teachers believe are the primary outcomes from the professional development they received?** | | | | | | | |
| **a. How has their teaching changed?** | | | | ◆ | ◆ | | |
| **b. How has student learning changed?** | | | ◆ | ◆ | ◆ | | |
| 3. **How could professional development for teachers be improved?** | | ◆ | ◆ | ◆ | ◆ | | |
| **Professional Development for School Literacy Coaches** | | | | | | | |
| 4. **To what extent did school literacy coaches participate in group professional development?** | | | | | | | |
| **a. How many hours of professional development were offered?** | | ◆ | | | | | ◆ |
| **b. What topics were offered?** | ◆ | ◆ | | | | | ◆ |
| **c. What percent of literacy coaches participated in each phase of professional development?** | | ◆ | | | | | ◆ |
| **d. To what extent were school literacy coaches satisfied with the group professional development they received?** | | ◆ | ◆ | ◆ | | | |
| 5. **How could professional development for school literacy coaches be improved?** | ◆ | ◆ | ◆ | ◆ | | | |
| 6. **To what extent did school literacy coaches participate in individual professional development with either University of Kansas or district staff?** | ◆ | | ◆ | ◆ | | | |

*(exhibit continues)*

**Exhibit 6 (continued)**

| Research Question | Surveys and Interviews: Developer | District Staff | Principal | Coach | Teacher | Classroom Observation | PPS Records Review |
|---|---|---|---|---|---|---|---|
| **Professional Development for School Administrators** | | | | | | | |
| 7. **To what extent did school administrators participate in group professional development?** | | | | | | | |
| **a. How many hours of professional development were offered?** | | ◆ | | | | | ◆ |
| **b. What topics were offered?** | | ◆ | | | | | ◆ |
| **c. What percent of school administrators participated in each phase of professional development?** | | ◆ | | | | | ◆ |
| **d. To what extent were school administrators satisfied with the group professional development they received?** | | ◆ | ◆ | | | | |
| 8. **How could professional development for school administrators be improved?** | | ◆ | ◆ | ◆ | | | |
| **Professional Development Summary** | | | | | | | |
| 9. **What factors contributed to the successful implementation of professional development?** | | ◆ | | | | | |
| **Instructional Coaching** | | | | | | | |
| 10. **To what extent did teachers participate in coaching?** | | | | | | | |
| **a. What percent of teachers worked with their school literacy coach?** | | | | ◆ | ◆ | | |
| **b. What activities did teachers work on with their school literacy coach?** | | | | ◆ | ◆ | | |
| **c. How much time did teachers spend, on average, with their school literacy coach?** | | | | ◆ | ◆ | | |
| **d. To what extent were teachers satisfied with the coaching they received from their school literacy coach?** | | | | ◆ | ◆ | | |
| **e. What percent of teachers worked with the University of Kansas professional developers?** | ◆ | | | | ◆ | | |
| **f. What topics did teachers work on with the University of Kansas professional developers?** | ◆ | | | | ◆ | | |
| **g. How much time did teachers spend, on average, with the University of Kansas professional developers?** | ◆ | | | | ◆ | | |
| **h. To what extent were teachers satisfied with the coaching they received from the University of Kansas professional developers?** | ◆ | | | ◆ | ◆ | | |

*(exhibit continues)*

**Exhibit 6 (continued)**

| Research Question | Surveys and Interviews: Developer | District Staff | Principal | Coach | Teacher | Classroom Observation | PPS Records Review |
|---|---|---|---|---|---|---|---|
| 11. **To what extent did school literacy coaches implement the specified instructional coaching activities?** | ◆ | | | ◆ | | | |
| 12. **How does coaching vary with teacher content area?** | | | | ◆ | | | |
| 13. **How does coaching vary for content and Xtreme teachers?** | | | | ◆ | | | |
| District- and School-Level Administrative Support | | | | | | | |
| 14. **What types of support has the district office provided to Striving Readers schools?** | | ◆ | ◆ | ◆ | ◆ | | |
| 15. **What additional support is needed from the district office?** | | ◆ | ◆ | ◆ | ◆ | | |
| 16. **What types of support have administrators provided to Striving Readers teachers and coaches?** | | | ◆ | ◆ | ◆ | | |
| 17. **What additional support is needed from school administrators?** | | ◆ | | ◆ | ◆ | | |
| Staff Characteristics | | | | | | | |
| 18. **What recruitment strategies were used to find teachers and coaches for Striving Readers?** | | ◆ | ◆ | | | | |
| 19. **What training and experience do teachers and coaches have?** | | | | | ◆ | | |
| 20. **To what extent did teacher and coach characteristics meet desired criteria?** | | ◆ | ◆ | | | | |
| Classroom Materials and Equipment | | | | | | | |
| 21. **What materials, software, and equipment were content teachers expected to have?** | | ◆ | | | ◆ | | |
| 22. **To what extent did content teachers implement the program with fidelity?** | | | | | | | |
| 23. **In what ways do content teacher materials, software, and equipment need to be improved?** | | ◆ | | ◆ | ◆ | | |
| 24. **How useful are the materials, software, and equipment for implementing the program?** | | ◆ | | ◆ | ◆ | | |

*(exhibit continues)*

**Exhibit 6 (continued)**

| Research Question | Surveys and Interviews: Developer | District Staff | Principal | Coach | Teacher | Classroom Observation | PPS Records Review |
|---|---|---|---|---|---|---|---|
| 25. **Instructional Processes** | | | | | | | |
| **a. How frequently do content teachers use the University of Kansas Content Enhancement Routines?** | | | | | ◆ | | ◆ |
| **b. What proportion of content teachers met frequency of use criteria?** | | | | ◆ | | | |
| **c. To what extent did content teachers implement the Content Enhancement Routines with fidelity?** | ◆ | | | ◆ | ◆ | ◆ | |
| **d. What were the barriers to successful implementation of the Content Enhancement Routines?** | ◆ | ◆ | ◆ | | ◆ | | |
| **e. What steps are University of Kansas and district staff taking to improve fidelity of program implementation?** | | ◆ | | | | | |
| **f. How do teachers rate the quality and effectiveness of the Strategic Instruction Model materials?** | | | | | ◆ | | |
| Implementation Summary | | | | | | | |
| 26. **Overall, in what important ways did the program as implemented differ from the program as planned?** | | ◆ | | | | | |
| 27. **To what extent did fidelity of implementation differ across schools, teachers, and years?** | ◆ | ◆ | ◆ | | ◆ | | |
| 28. **What is the overall level of satisfaction with the Striving Readers program?** | ◆ | ◆ | ◆ | ◆ | ◆ | | |
| Impact of Whole School Intervention | | | | | | | |
| 29. **To what extent do students in Grades 6–12 improve their reading scores on the OAKS after implementation of the University of Kansas content enhancement routines?** | | | | | | | |

# Implementation of the Targeted Intervention

To provide context for interpreting the impact of the targeted intervention, the evaluation team assessed the implementation of the Xtreme Reading curriculum in Years 1 through 5. Program staff expected implementation to improve as teachers gained experience using the curriculum. Although many factors contributed to the fidelity of implementation, barriers to implementation also emerged. This chapter describes the implementation evaluation design, the findings from Years 1 through 5, and the implications of these findings for the impact analyses.

## Implementation Evaluation Design

The evaluation team collected data for the targeted intervention implementation evaluation by surveying or interviewing the professional developers, the Striving Readers district leadership team, school administrators, school literacy coaches, and Xtreme Reading teachers and by conducting classroom observations. The data were analyzed in terms of professional development inputs, classroom implementation fidelity, and teacher buy-in.

- **Professional development inputs** were defined as teachers' (a) group professional development attendance and (b) qualifications. Professional development offered in Year 1 included a summer training session, fall make-up training sessions, and periodic workshops. Professional development offered in Year 2 included a summer training session and fall make-up training sessions. Professional development offered in Years 3 through 5 included a summer training session and monthly Xtreme Reading teacher meetings. If more than one teacher taught the targeted intervention at a school, their professional development attendance was averaged for a school-level score.[1] In terms of qualifications, the district's goal was for at least one Xtreme Reading teacher at each school to have a reading endorsement.[2]
- **Classroom implementation fidelity** was defined as the average of the fidelity ratings that were part of the classroom observations conducted by the evaluation team. During Year 1 the fidelity ratings were derived from one classroom observation conducted in winter and another conducted in spring 2007. In Years 2 through 5 the fidelity ratings

[1]Professional development attendance data provided by Portland Public Schools.
[2]Teachers self-reported their qualifications on the teacher survey.

were derived from between 3 and 7 classroom observations conducted throughout the academic year.

- **Teacher buy-in** was indicated by the teachers' survey responses in 2 domains: (a) their perceptions of the group professional development and in-school coaching provided by the professional developers as measured by their agreement on a 5-point scale with statements such as "The professional development on the Xtreme Reading program prepared me to use the program effectively in my classroom" and "Observations conducted by the professional developers have helped me to implement the Xtreme Reading program," and (b) their perceptions of the effectiveness of the Xtreme Reading strategies (see Exhibit 2) as measured by their ratings on a 5-point scale. The responses in the first domain were averaged across the fall, winter, and spring in Year 1; across the fall and winter in Year 2; and across the fall and spring in Years 3 through 5. The responses in the second domain were averaged across the winter and spring in Year 1; across the fall, winter, and spring in Year 2; and across the fall and spring in Years 3 through 5.

## Implementation Evaluation Findings

Exhibit 7 summarizes the levels of implementation of Xtreme Reading attained at each of the 10 veteran participating schools in Years 1 through 5. Exhibit 8 summarizes the levels of implementation for the 10 new schools that participated in Xtreme Reading in Year 5 only. (See Appendix B for details on the calculation of ratings for each area of implementation.)

### *Professional Development*

In terms of professional development inputs, in Year 1 40% of the middle school teachers and 25% of the high school teachers attended more than 75% of the group professional development sessions offered. In Year 2 these proportions increased: 89% of the middle school teachers and 75% of the high school teachers attended more than 75% of the professional development sessions. In Year 3 all of the middle school teachers and 75% of the high school teachers attended more than 75% of the professional development sessions. This pattern persisted in Year 4: 100% of middle school teachers and 75% of high school teachers attended more than 75% of the professional development sessions.

In Year 5, Xtreme Reading teachers with at least 2 years of experience (n = 8) were not required to attend professional development sessions. Teachers at the 10 new schools (n = 11)

and 4 new teachers at veteran schools were expected to attend summer training and attend 6 meetings during the school year; 80% of the new teachers attended at least 75% of the professional development sessions.

### *Teacher Qualifications*

In Year 1 only 3 of the 5 middle schools and 1 of the 4 high schools had an Xtreme Reading teacher with a reading endorsement. These proportions increased in Year 2 such that 4 of the middle schools and 3 of the high schools had an Xtreme Reading teacher with a reading endorsement. These numbers remained the same in Year 3, but changed again in Year 4: 5 of the 6 middle schools had teachers with a reading endorsement but only 1 of the 4 high schools did. In the final year of Striving Readers, 58% of the 19 middle school teachers had a reading endorsement as did 25% of the 4 high school teachers.

### *Classroom Implementation Fidelity*

Classroom implementation fidelity was generally high in the middle schools and medium in the high schools in Year 1 (with one low level school in each category). In Year 2 classroom implementation fidelity remained steady or declined among the middle schools but declined across all the high schools. In Year 3 classroom implementation fidelity improved in 5 schools, declined in 1 school, and stayed the same in 1 school. In Year 4 implementation fidelity stayed the same in 5 schools, declined in 3 schools, and improved in 2 schools. In Year 5 implementation fidelity stayed the same in 2 schools, declined in 5 schools, and improved in 3 of the veteran schools. These changes in observed fidelity occurred in the context of some teacher turnover but with similar involvement in professional development and teacher buy-in. Implementation fidelity at the 10 new middle schools ranged from very low to high.

### *Teacher Buy-In*

Teacher buy-in, as measured by teacher responses to survey questions concerning professional development and perceived effectiveness of the Xtreme Reading strategies, was generally higher at the middle schools than at the high schools in all years. Teacher buy-in remained relatively stable at both school levels across all years, generally ranging from medium to high among the middle schools and from low to high among the high schools. Teacher buy-in at the 10 new middle schools in Year 5 ranged from low to high.

## Exhibit 7
## Implementation of the Targeted Intervention

| | Professional Development Inputs | | | | | Classroom Implementation Fidelity | | | | | Teacher Buy-In | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **School** | **Yr 1** | **Yr 2** | **Yr 3** | **Yr 4** | **Yr 5** | **Yr 1** | **Yr 2** | **Yr 3** | **Yr 4** | **Yr 5** | **Yr 1** | **Yr 2** | **Yr 3** | **Yr 4** | **Yr 5** |
| **Middle** | | | | | | | | | | | | | | | |
| Alpha | H | H | H | H | H | H | M | L | M | H | M | H | H | H | H |
| Beta | M | H | H | H | H | H | H | H | H | H | M | M | H | H | H |
| Gamma | H | H | H | H | H | H | M | H | M | vL | H | H | H | H | M |
| Delta | M | H | H | H | H | L | M | H | H | M | M | M | H | M | L |
| Epsilon | H | H | — | — | — | H | H | — | — | — | H | M | — | — | — |
| Zeta | — | — | H | H | H | — | — | M | L | M | — | — | M | M | H |
| Omega | — | — | H | H | H | — | — | H | H | M | — | — | H | H | M |
| **High** | | | | | | | | | | | | | | | |
| Kappa | L | H | H | H | H | M | vL | H | L | M | L | M | H | M | H |
| Lambda | M | H | H | M | H | M | L | M | M | L | M | M | M | M | M |
| Sigma | H | H | M | H | H | M | vL | L | L | L | M | M | M | M | M |
| Theta | L | M | H | H | M | L | — | M | H | M | L | L | M | M | M |

*Note*. H = *high*, M = *medium*, L = *low, vL = very low*. In Year 2 Beta middle school and Delta middle school merged with K–5 schools; Epsilon middle school closed and the final Grade 8 class went to Kappa high school; Gamma middle school became an all-girls school with new Striving Readers teachers. In Year 3 Alpha middle school merged with a K–5 school and 2 new middle schools were added to Striving Readers: Zeta and Omega.
In Year 2 the Xtreme Reading teacher at Theta high school did not permit classroom observations.

**Exhibit 8**
**Implementation of the Target Intervention**
**for New Schools in Year 5**

| School | Professional Development Inputs | Classroom Implementation Fidelity | Teacher Buy-In |
|---|---|---|---|
| **A** | H | L | M |
| **B** | H | H | H |
| **C** | H | M | H |
| **D** | H | M | H |
| **E** | M | vL | H |
| **F** | M | M | H |
| **G** | H | M | H |
| **H** | M | H | L |
| **I** | H | H | L |
| **J** | H | L | H |

## Implications for the Impact Analyses

Interpretation of the impact of the targeted intervention on student reading achievement and motivation is facilitated by an examination of the issues that emerged during the 5 years of implementation of the curriculum. This section focuses on factors that facilitated the implementation of Xtreme Reading, differences in implementation across school levels, and barriers to successful implementation of the curriculum.

### *Factors That Facilitated Implementation*

Overall, in Year 1 the targeted intervention teachers were satisfied with the Xtreme Reading curriculum. The survey results indicated that the reading materials were of high interest to the students and the teachers experienced few difficulties motivating students to read. In most instances the teachers appreciated having a coteacher and access to a school literacy coach and professional developers when they needed assistance.

Several changes made during Year 2 served to facilitate implementation. First, the targeted intervention teachers were able to choose from 3 schedule options for implementing the Xtreme Reading curriculum. Whereas the original schedule clustered all of the goal setting and

classroom management units at the beginning of the school year, the new schedule options spread out these units and introduced the academic units earlier. Other changes included incorporating more age-appropriate books into the middle school curriculum and allowing coteaching to be optional. The professional developers also began implementing monthly meetings for the Xtreme Reading teachers, which afforded them additional opportunities to receive training on specific aspects of the curriculum and share effective instructional strategies. In addition, the school literacy coaches began working with both the targeted intervention teachers and the content teachers.

Satisfaction with the Xtreme Reading curriculum remained high in Year 2, and the targeted intervention teachers gained confidence in their skills. The survey results indicated that the teachers were most satisfied with the feasibility of applying reading strategies to other classes, the clarity of purpose for each lesson, the amount of repetition and review across lessons, and the appropriateness of the reading subject matter for their students. The targeted intervention teachers reported that the monthly meetings with other Xtreme Reading teachers facilitated implementation of the curriculum. They also expressed appreciation for the assistance they received from the professional developers and the school literacy coaches.

In Year 3, 5 of the 13 (38%) Xtreme Reading teachers were new, and teachers reported that the support from University of Kansas (KU) trainers and district staff was very helpful—especially the modeling of the strategies in the classroom. Teachers also believed that the program improved students' reading skills and attitudes toward reading. They cited the availability of multiple book sets and allowing students to choose the texts for book study activities as particularly helpful.

In Year 4, 3 of the 11 (27%) Xtreme Reading teachers were new. Teachers continued to report high levels of satisfaction with the professional development, especially the monthly meetings with other Xtreme Reading teachers. The majority of the Xtreme Reading teachers (70%) reported receiving from 1 to 5 hours of individual assistance from the KU Xtreme Reading expert, but 20% said they received over 6 hours of individual assistance. Eighty percent also reported receiving at least 1–5 hours of individual assistance from the district's Striving Readers Leadership Team. Most teachers reported that they received some help from their school's literacy coach or their school's KU professional developer as well. Teachers also appreciated the small class sizes that helped them to build strong relationships with students and the success their students achieved as a result of learning new reading strategies.

In Year 5, all of the 11 teachers at the 10 new K–8 or middle schools were new to Xtreme Reading and 4 of the 12 teachers at the 10 veteran Striving Readers schools were new. Both new and experienced teachers gave high ratings to the professional development, especially for demonstration lessons and meetings with other Xtreme Reading teachers. Most teachers liked the reading strategies taught in the program, especially paraphrasing, visual imagery, and self questioning. Several teachers expressed satisfaction with the growth they saw in student reading skills, the transfer of those skills to other classes, small class sizes, and the book library that came with the program.

### *Variation in Implementation by School Level*

In Years 1 through 3 the middle school teachers participated in more professional development than the high school teachers. For example, 90% of the middle schools had teachers who participated in more than half of the sessions offered in Year 1, compared to 62% of the high schools. In Year 2 all of the middle schools had teachers who participated in more than half of the professional development sessions, compared to 75% of the high schools. In Year 3, all middle school teachers participated in more than half of the professional development sessions, compared to 71% of the high school teachers. By Year 4, however, 100% of middle school teachers and 100% of high school teachers attended at least half of the professional development activities. In Year 5, only the 15 new teachers were required to participate in professional development, and most of these teachers (n = 13) taught at the middle school level. All of the teachers attended at least half of the professional development activities.

With regard to teacher qualifications, in Year 1 60% of the middle schools had an Xtreme Reading teacher with a reading endorsement compared to only 25% of the high schools. In Year 2 these percentages increased to 80% of the middle schools and 75% of the high schools. In Year 3, 67% of the middle schools and 75% of the high schools had an Xtreme Reading teacher with a reading endorsement. In Year 4 86% of the middle school teachers had a reading endorsement but only 25% of the high school teachers did. In Year 5 67% of the middle school teachers had a reading endorsement compared to 25% of the high school teachers.

Given the rigorous, prescribed curriculum of the targeted intervention, variability in professional development inputs across school levels could have implications for fidelity of implementation, thereby impacting student outcomes. In general, classroom implementation fidelity and teacher buy-in were higher at the middle school level than the high school level in all years. In addition,

attendance rates at the targeted middle schools (91% in Years 1 and 2, 93% in Years 3 and 4, and 94% in Year 5) were higher than those at the targeted high schools (85% in Year 1, 87% in Year 2, 86% in Years 3 and 4, and 87% in Year 5).

### *Barriers to Implementation*

The Xtreme Reading teachers were surveyed in the fall, winter, and spring of Years 1 and 2 and in the fall and spring of Years 3 through 5. In addition, the study team interviewed each Xtreme Reading teacher in the spring of each year. Teachers were asked to comment on the program components that they were most satisfied with and to suggest improvements to the Striving Readers program. In all years the majority of the teachers reported that the curriculum was well organized and easy to use, although several respondents did express concerns about certain aspects of the curriculum. In Year 1 some teachers considered the behavioral units too juvenile for high school students, and some believed that too much time was allocated to the behavior units at the beginning of the year. Teachers also considered the reading content too mature for middle school students and some teachers described the book study component as poorly developed. Suggestions included providing a wider variety of reading materials, using page numbers and color coding to organize the student binders, and providing more trade books for the classroom libraries.

In Year 2 teachers expressed dissatisfaction with the independent and paired practice reading materials—in particular the materials in the Self-Questioning and Visual Imagery units. They requested selections that would be more engaging and relevant for students. Teachers also suggested reducing the amount of content they are expected to cover, although they came closer to completing the curriculum in Years 2 through 5 than in Year 1. For example, in Year 1 none of the teachers completed the Paraphrasing unit or started the Inference unit, but in Years 2 through 5 most completed Paraphrasing and started Inference. In Year 3 the Xtreme Reading teachers suggested that the teacher and student notebooks could be better organized and more compact and should correspond more closely with each other. Teachers also asked for more current and interesting leveled readings within the student materials and a better definition of what the levels mean. In the absence of developer revisions to the program, these criticisms were repeated in Years 4 and 5.

District staff had expected to receive a completed Xtreme Reading curriculum in Year 1 and were disappointed that the curriculum was still under development. District staff were dismayed

that the curriculum materials they received required extensive duplication and organization prior to distribution, which diminished the time available to provide direct assistance to teachers and school literacy coaches. This problem was reduced somewhat in Year 2 because new materials were available earlier. In Year 3 district staff complained of too many notebooks of materials and observed that teachers were confused about which vocabulary words to use because the words in the notebooks did not correspond to the reading materials. District staff also noted that the unit pre- and posttests were of poor quality and the oral reading passages were outdated and uninteresting to students. Some of these concerns were addressed by district Striving Readers staff in Years 4 and 5 through improvements in the organization of student notebooks, the provision of additional reading passages for student practice, and additional classroom library books.

Staffing issues also presented challenges to the implementation of Xtreme Reading. Whereas in Year 1 all Xtreme Reading classrooms were staffed by a pair of coteachers, in Year 2 this proportion decreased to 67% of classrooms and in Years 3 through 5 all Xtreme Reading classrooms were staffed by a single teacher. Implementation barriers related to coteaching included a lack of time to collaborate and coordinate (both within and across schools); inadequate training on coteaching strategies; and, in some cases, a lack of careful screening to ensure the compatibility of the coteachers. Another important staffing issue in Year 1 was the delay in hiring 6 of the 9 Xtreme Reading content teachers, which resulted in their missing the professional development conducted in August 2006. In Year 1 Xtreme Reading teachers expressed concern that school administrators did not fully understand and support the Striving Readers program, and the language skills of some students selected for the intervention were either too high or too low. According to the professional developers, some Xtreme Reading teachers struggled to implement the curriculum with fidelity and demonstrated a need for additional training with an emphasis on following lesson plans and pacing. Many of the targeted intervention teachers also experienced significant challenges managing student behavior.

In Year 2 changes in school configuration impeded implementation of the Striving Readers program. District administrators closed one Striving Readers middle school, converted 2 Grades 6–8 middle schools to Grades K–8 schools, and converted a coeducational Grades 6–8 middle school to an all-girls school serving Grades 6–9. These changes affected enrollment and reduced the number of students eligible for the targeted intervention. Nevertheless, only 2 new Xtreme Reading teachers had to be hired for Year 2. In Year 3, two additional Grades 6–8 middle schools were added to the Striving Readers program and a Grade 8 academy on one of

the high school campuses closed, bringing the total number of participating schools to 10. These and other staffing changes resulted in 5 new Xtreme Reading teachers to be trained in Year 3. In Year 4, only 3 new Xtreme Reading teachers needed training. However, in Year 5 the district decided to drop the experimental design because program effectiveness had already been demonstrated in Years 1 through 4 and the district wanted to be able to serve all students who qualified for the program. In addition, the district decided to direct remaining resources to the expansion of Xtreme Reading into 10 new K–8 and middle schools. Thus, in the final year of Striving Readers, district staff focused on training 15 new Xtreme Reading teachers (11 new teachers for the new schools and 4 new teachers for the veteran schools).

The professional developers estimated spending about half of their time in Year 1 and 15% to 80% of their time in Year 2 providing direct assistance to the targeted intervention teachers. They remained concerned about the fidelity of implementation in some Xtreme Reading classrooms, but overall the professional developers were pleased with the implementation improvements observed in Year 2. In Year 3, the district leadership team assumed greater responsibility for supporting the Xtreme Reading teachers through monthly meetings and class visits. Because 5 of the teachers were new, there was a wide range of experience and skill in the group, which posed some challenges for professional development. In Year 4, the district leadership team and the KU Xtreme Reading expert continued to work together to support the Xtreme Reading teachers, but responsibility for the training continued to shift more toward the PPS staff. This change was intentional in order to increase district capacity for supporting the Xtreme Reading program. By Year 5, all of the professional development for new teachers was provided by PPS staff.

In summary, although teachers experienced difficulties implementing the Xtreme Reading program, many of the issues were addressed by district staff or the professional developers over the 5 years of the grant. Most teachers reported that Xtreme Reading was effective with struggling readers and believed they had received adequate training and support to implement the program. The middle school teachers did, however, tend to have better professional development attendance, better classroom implementation fidelity, and higher levels of buy-in than the high school teachers. The next section of this report shows that these differences are reflected in the impact of the program on middle and high school students.

# Impact of the Targeted Intervention

The evaluation team assessed the impact of Xtreme Reading in Years 1 through 4. The impact evaluation examined differences between the treatment and control groups in terms of reading achievement on the GRADE and the OAKS and student reading motivation. Student outcome data for Year 5 are not included in this analysis because the district elected to discontinue the experimental design for the final year of funding. Instead, the district funded 9 new K-8 schools and 1 new middle school to implement the Xtreme Reading program, and all 21 Striving Readers schools were permitted to serve any student who qualified for the program. This section of the report describes the impact evaluation design, the characteristics of the students included in the evaluation, and the findings from the data analysis.

## Impact Evaluation Design

The targeted intervention impact evaluation addressed 5 evaluation questions:

- To what extent do students in Xtreme Reading improve their reading skills after one year compared to the control group?
- Do the effects of Xtreme Reading differ by school level?
- Do the effects of Xtreme Reading differ by cohort?
- What percentage of students in Xtreme Reading are reading at grade level by the end of the school year?
- To what extent do students in Xtreme Reading improve their reading motivation after one year compared to the control group?

This section describes the impact evaluation sampling plan, the counterfactual for Xtreme Reading, the data collection plan, and the analytic approach used to address the impact evaluation questions.

### *Sampling Plan*

Students in Grades 7–10 are eligible for the targeted intervention if they are reading at least 2 years below grade level as measured by their total reading score on the OAKS or the Degrees of Reading Power (for Grade 10 selection only) administered the previous spring. Prior to random assignment, students are screened to ensure that participation in the targeted

intervention does not conflict with special education requirements or English language learner requirements.[3] To provide schools with sufficient time for scheduling students into classes, most targeted intervention students are identified and randomly assigned in the spring. When school begins in the fall, potential targeted intervention participants with no OAKS or Degrees of Reading Power scores are administered the GRADE to determine eligibility for random assignment.

In each of the schools participating in Striving Readers, between 18% and 38% of the students were reading 2 or more years below grade level on the OAKS prior to the implementation of the program. On average, 27% of the schools' student population were deemed eligible for the targeted intervention. The number of students who met the eligibility criteria for Striving Readers and were randomly assigned across Years 1, 2, 3, and 4 included a combined total of 2,895 students. Of those, 1,453 students were assigned to the treatment group and 1,442 to the control group.

In Year 1 a total of 1,311 students were randomly assigned to the treatment group (*n* = 659) or the control group (*n* = 652). Of the 1,311 randomly assigned students, 348 treatment and 427 control students received the allocated intervention or control group condition as planned. Of the 311 treatment group students and 225 control group students who did not receive the allocated intervention or control group conditions as planned, 62 treatment and 39 control group students were excluded because of individualized education plan (IEP) conflicts that made them ineligible for both the intervention and study. An additional 115 treatment and 129 control group students were excluded from the analytic target sample because they never enrolled in a Striving Readers school. Thus in Year 1 a total of 177 treatment and 168 control group students were ineligible for both the intervention and the study. Reasons that treatment group students were exempt from the intervention included reading skills that were too high,[4] parent withdrawal, scheduling conflicts, or transfer to another school. In addition, 9 control group students—crossovers—received the intervention and thus did not receive the allocated control.

[3]In Year 1 screening for special education and English learner needs that might conflict with participation in Xtreme Reading did not occur until after random assignment. In Years 2 and 3 a few students were missed in the screening process and were determined after random assignment occurred to have special education and English learner needs that conflicted with Xtreme Reading.

[4]Using student course grades or other test scores, school staff determined that some students were inappropriately identified for the intervention and were removed from the study after random assignment took place. Staff appeared to be more diligent in making this determination for treatment group students than control group students.

In Year 2 a total of 443 students were randomly assigned to the treatment group (*n* = 219) or the control group (*n* = 224). Of the 443 randomly assigned students, 134 treatment and 162 control students received the allocated intervention or control group condition. Of the 85 treatment group students and 62 control group students who did not receive the intervention or control group condition, 14 treatment and 4 control group students were excluded because of IEP conflicts that made them ineligible for both the intervention and the study, and 51 treatment and 39 control group students were excluded because they never enrolled in a Striving Readers school. Thus the analytic target sample in Year 2 included 154 treatment and 181 control group students.

In Year 3 a total of 684 students were randomly assigned to the treatment group (*n* = 345) or the control group (*n* = 339). Of the 684 randomly assigned students, 239 treatment and 253 control students received the allocated intervention or control group condition. Of the 110 treatment group students and 86 control group students who did not receive the intervention or control group condition, 10 treatment and 1 control group students were excluded because of IEP conflicts that made them ineligible for both the intervention and the study, and 44 treatment and 55 control group students were excluded because they never enrolled in a Striving Readers school. The analytic target sample in Year 3 included 291 treatment and 283 control group students.

In Year 4 a total of 457 students were randomly assigned to the treatment group (*n* = 230) or the control group (*n* = 227). Of those students who were randomly assigned, 168 treatment and 196 control students received the allocated intervention or control group condition. Of those who did not receive the allocated intervention or control group condition, 6 treatment students were excluded because of IEP conflicts that made them ineligible for both the intervention and the study, and 31 treatment and 26 control group students were excluded because they never enrolled in a Striving Readers school. The Year 4 analytic target sample included a total of 193 treatment students and 201 control group students.

### GRADE Analytic Sample

The analytic target sample included all students who were randomly assigned and enrolled in a Striving Readers school and did not have IEP conflicts. The combined analytic sample for Years 1 through 4 included 1,579 students (treatment group *n* = 756; control group *n* = 823). (See Appendix C for a more detailed breakdown of the combined GRADE analytic sample by

school level.) In Year 1 the analytic target sample *n*s were 482 and 484 for the treatment and control groups, respectively. Of the 482 treatment group students in the analytic target sample, 296 students had posttest GRADE scores and were included in the final analysis. Of the 484 control group students in the analytic target sample, 334 students had posttest GRADE scores and were included in the analysis. Exhibit 9 provides a detailed breakdown of the final GRADE analytic sample for Year 1.

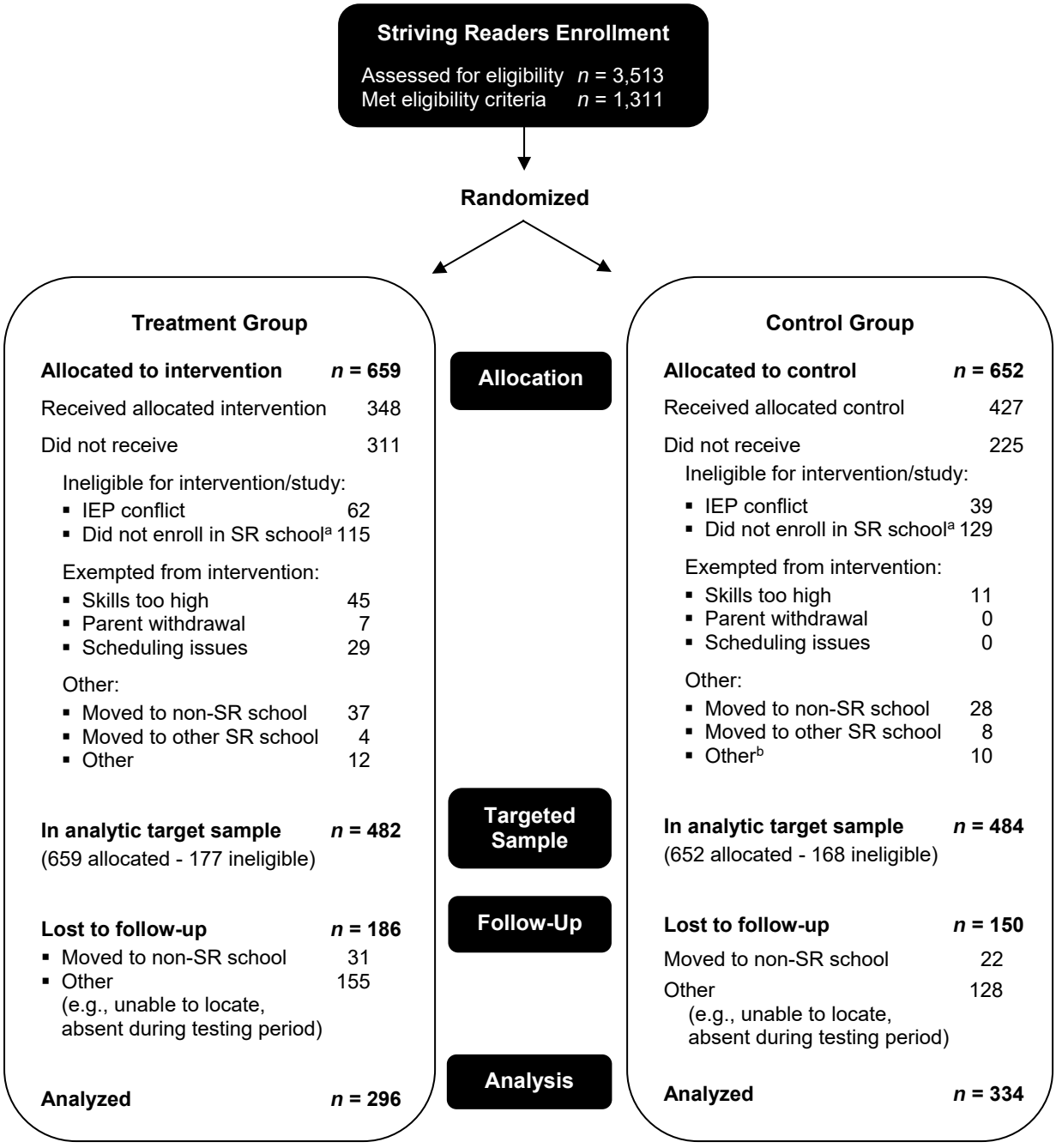
In Year 2 the analytic target sample included 154 treatment and 181 control group students. Of the 154 treatment group students in the analytic target sample, 87 students had posttest GRADE scores and were included in the final analysis. Of the 181 control group students in the analytic target sample, 132 students had posttest GRADE scores and were included in the final analysis. Exhibit 10 provides a detailed breakdown of the final GRADE analytic sample for Year 2.

The analytic target sample for Year 3 included 291 treatment and 283 control group students. Of these 216 treatment group and 208 control group students had posttest GRADE scores and were included in the final analysis. Exhibit 11 provides a detailed breakdown of the final GRADE analytic sample for Year 3.

The analytic target sample for Year 4 included 193 treatment and 201 control group students. Of these, 157 treatment group students and 149 control group students had posttest GRADE scores and were included in the final analysis. Exhibit 12 provides a detailed breakdown of the final GRADE analytic sample for Year 4.

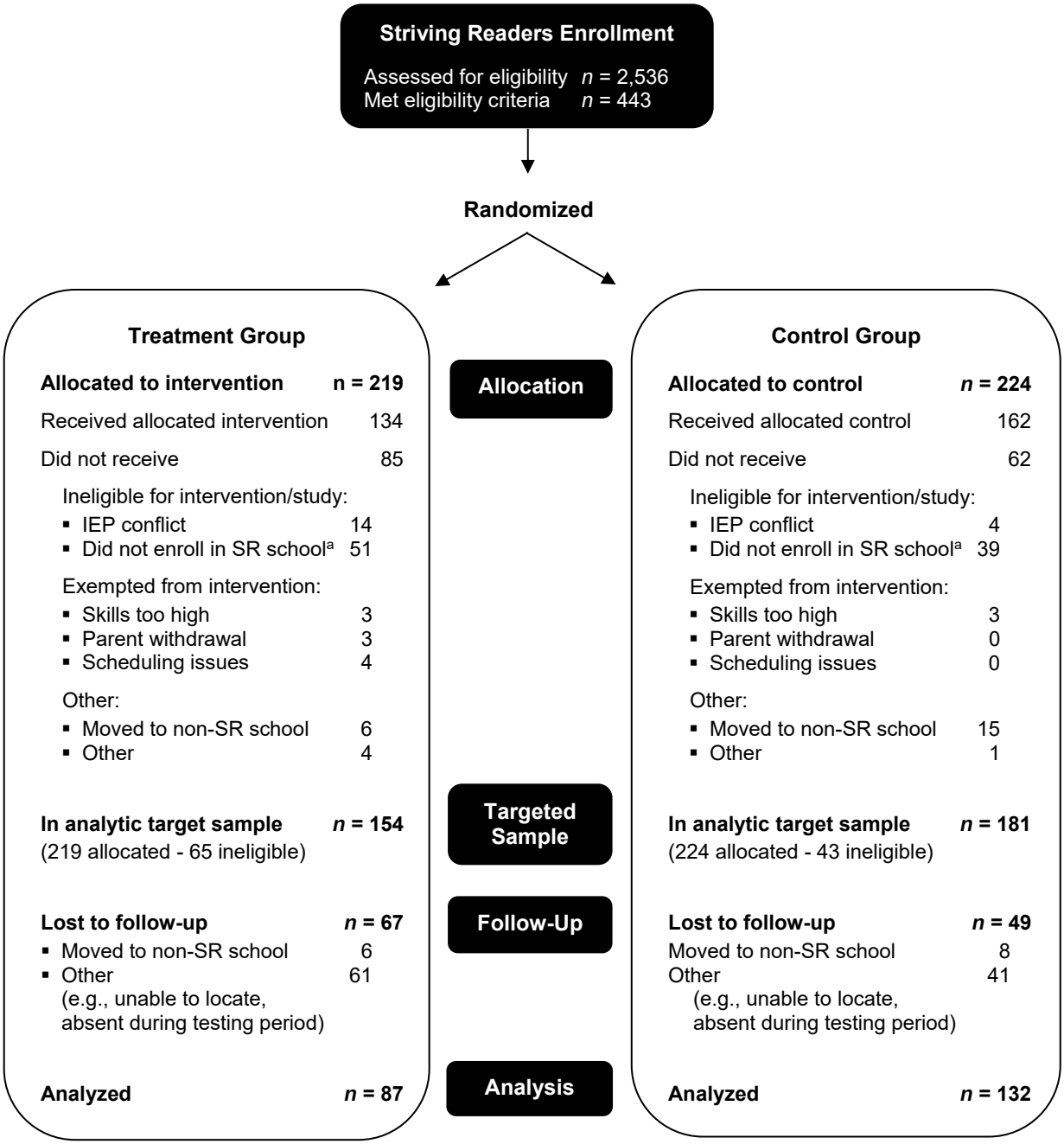
To assess whether selective study attrition occurred in the GRADE analytic sample, the evaluation team conducted equivalence tests on baseline demographic characteristics and OAKS reading scores. There were no differences on baseline OAKS reading scores between students in the analytic sample and students who were not. Students in the final analytic sample were, however, more likely than students not included in the sample to have an IEP and to be English language learners. Students in the final analytic sample also were more likely than students not included in the sample to be Hispanic and less likely to be African American. Appendix D provides further baseline equivalence test results for the GRADE analytic sample by school level. Given an assumed power of .80 and a 0.05 level of statistical significance for a 2-tailed test, the minimum detectable effect size for a within-school randomized trial model and a sample size of 1,579 is .10 (based on *Optimal Design* software).

**Exhibit 9**
**Targeted Intervention Random Assignment: Year 1 GRADE Sample**



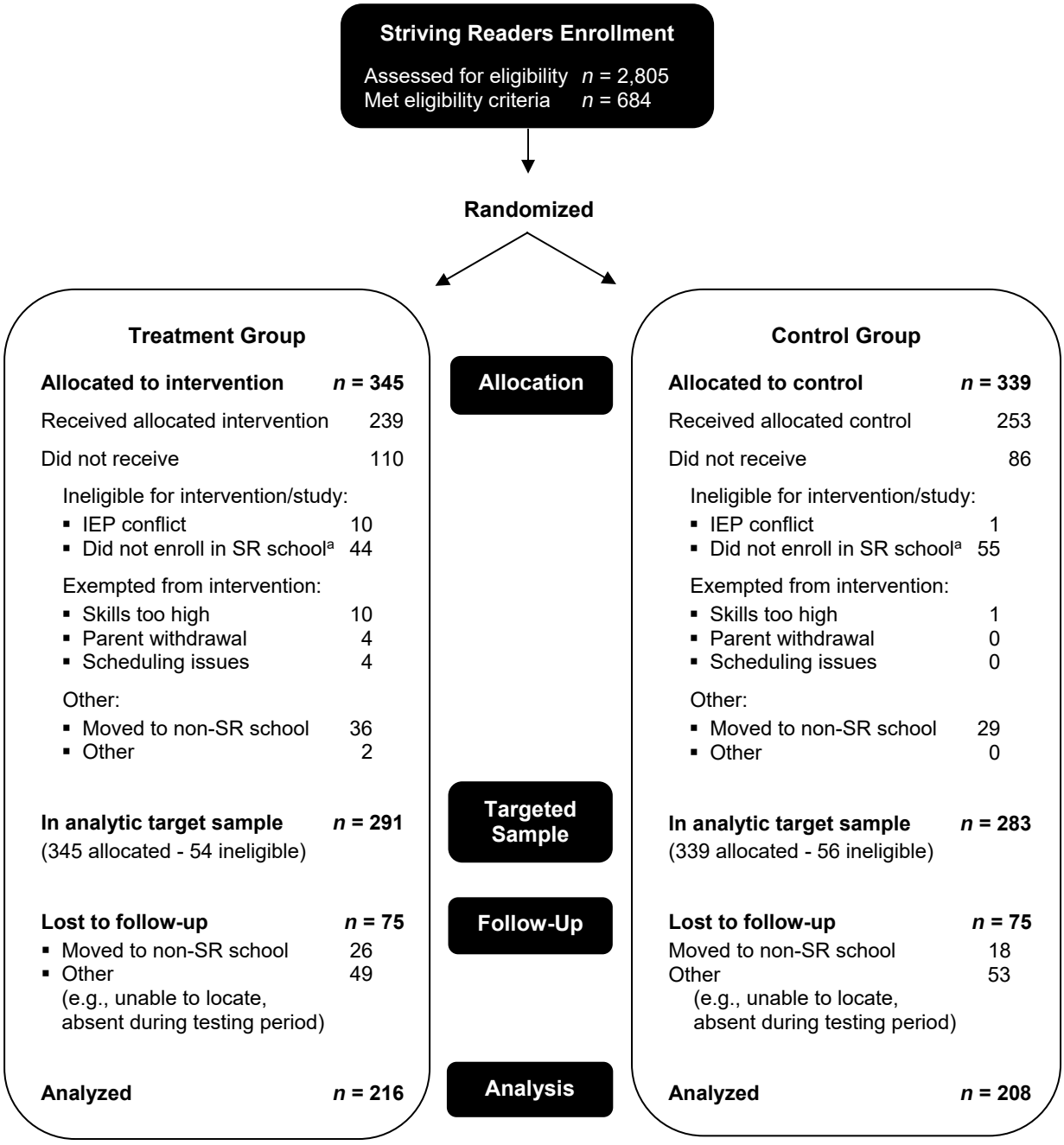

*Note.* IEP = individualized education program. SR = Striving Readers.
[a]These students never learned their group assignment. [b]Crossovers = 9.

**Exhibit 10**
**Targeted Intervention Random Assignment: Year 2 GRADE Sample**



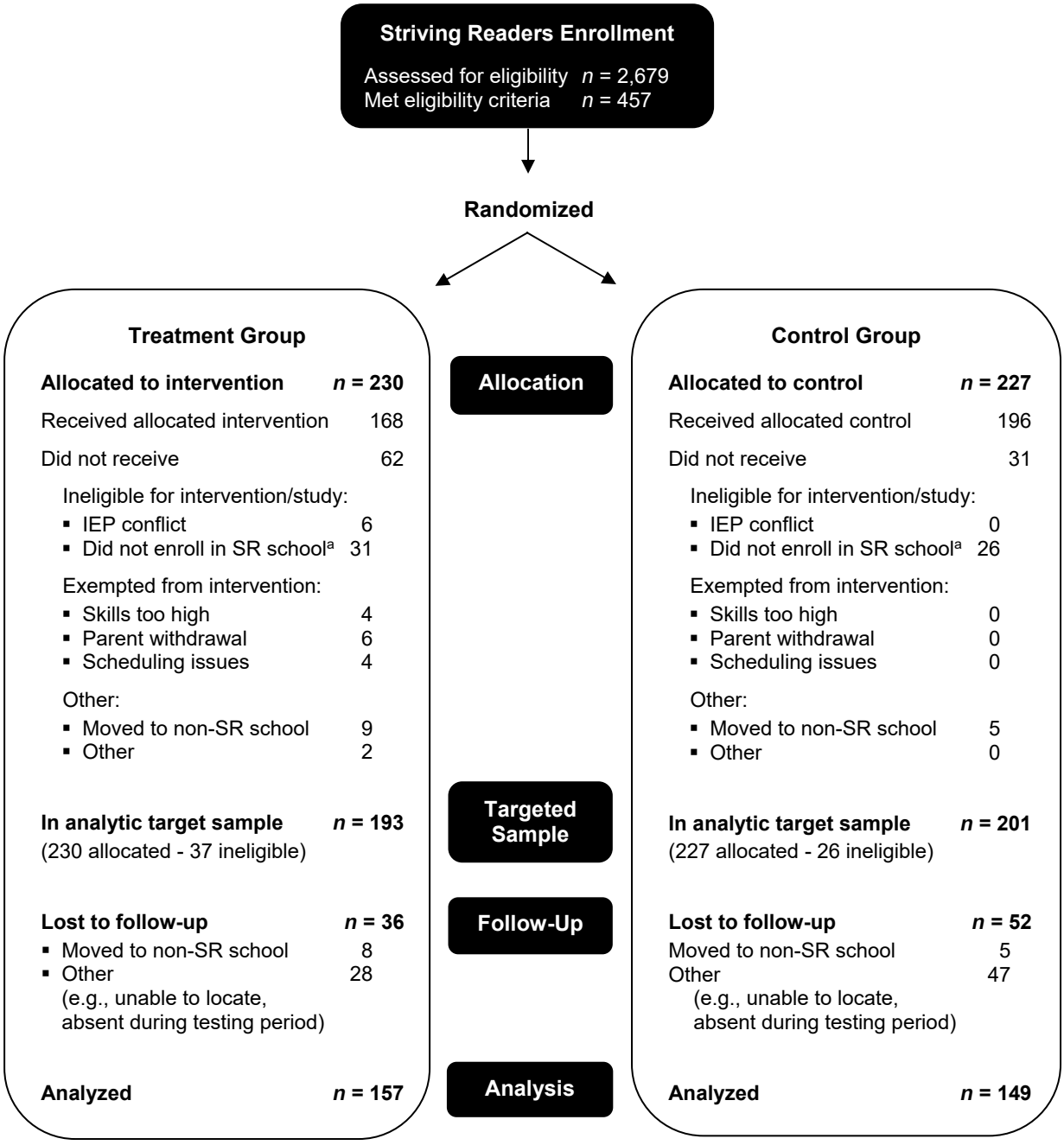

*Note.* IEP = individualized education program. SR = Striving Readers.
[a]These students never learned their group assignment.

**Exhibit 11**
**Targeted Intervention Random Assignment: Year 3 GRADE Sample**




*Note.* IEP = individualized education program. SR = Striving Readers.
[a]These students never learned their group assignment.

**Exhibit 12**
**Targeted Intervention Random Assignment: Year 4 GRADE Sample**




*Note.* IEP = individualized education program. SR = Striving Readers.
[a]These students never learned their group assignment.

### OAKS Analytic Sample

The OAKS analytic target sample included all students who were randomly assigned and enrolled in a Striving Readers school, did not have IEP conflicts, and were in a grade that administered the OAKS. The combined analytic sample for Years 1 through 4 included 1,468 students (treatment group *n* = 732; control group *n* = 736). (See Appendix E for a more detailed breakdown of the combined OAKS analytic sample by school level.) In Year 1 the OAKS analytic target sample *n*s were 407 for both the treatment and control groups. Of the 407 treatment group students in the analytic target sample, 322 had posttest OAKS scores and were included in the final OAKS analytic sample. Of the 407 control group students in the analytic target sample, 330 had posttest scores and were included in the final analysis. Exhibit 13 provides a detailed breakdown of the final OAKS analytic sample for Year 1.

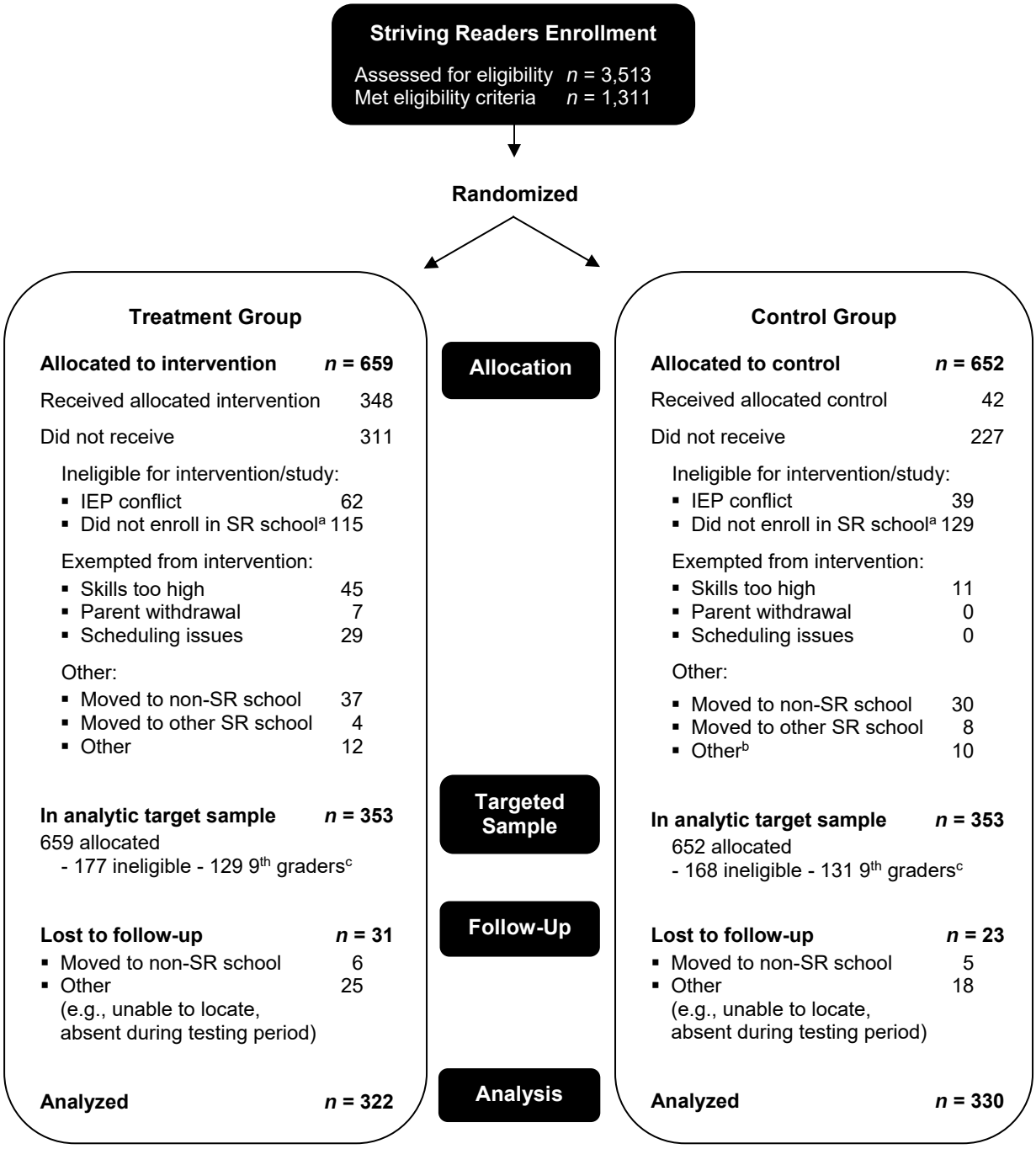
The analytic target sample for Year 2 included all students who were randomly assigned and enrolled in a Striving Readers school and did not have IEP conflicts (*n*s = 131 and 135 for the treatment and control groups, respectively). Of the 131 treatment group students in the analytic target sample, 105 had posttest scores and were included in the analysis. Of the 135 control group students in the analytic target sample, 103 had posttest scores and were included in the analysis. Exhibit 14 provides a detailed breakdown of the final OAKS analytic sample for Year 2.

The analytic target sample for Year 3 included 211 treatment group and 206 control group students. Of these, 184 treatment and 181 control group students had posttest scores and were included in the analysis. Exhibit 15 provides a detailed breakdown of the final OAKS analytic sample for Year 3.

The analytic target sample for Year 4 included 128 treatment and 132 control group students. Of these, 121 treatment and 122 control group students had posttest scores and were included in the analysis. Exhibit 16 provides a breakdown of the final OAKS analytic sample for Year 4.

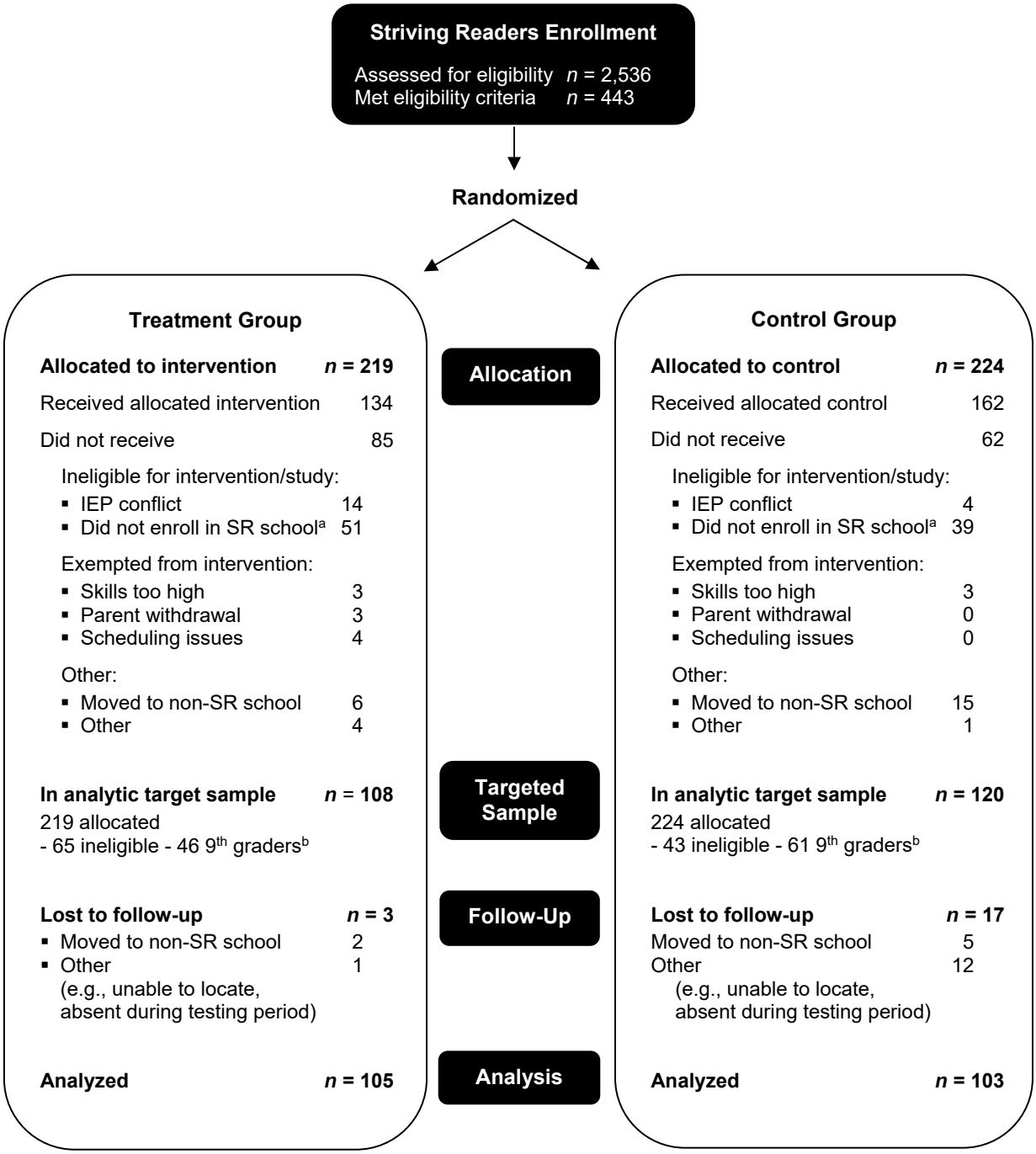
To assess whether selective study attrition occurred in the OAKS analytic sample, the evaluation team conducted equivalence tests on baseline demographic characteristics and OAKS reading scores. There were no differences on baseline OAKS reading scores between students in the analytic sample and students who were not. However, students in the final analytic sample were more likely than students not included in the sample to be English language learners and Hispanic. Baseline equivalence results for the OAKS analytic sample are broken out by school level in Appendix D.

**Exhibit 13**
**Targeted Intervention Random Assignment: Year 1 OAKS Sample**



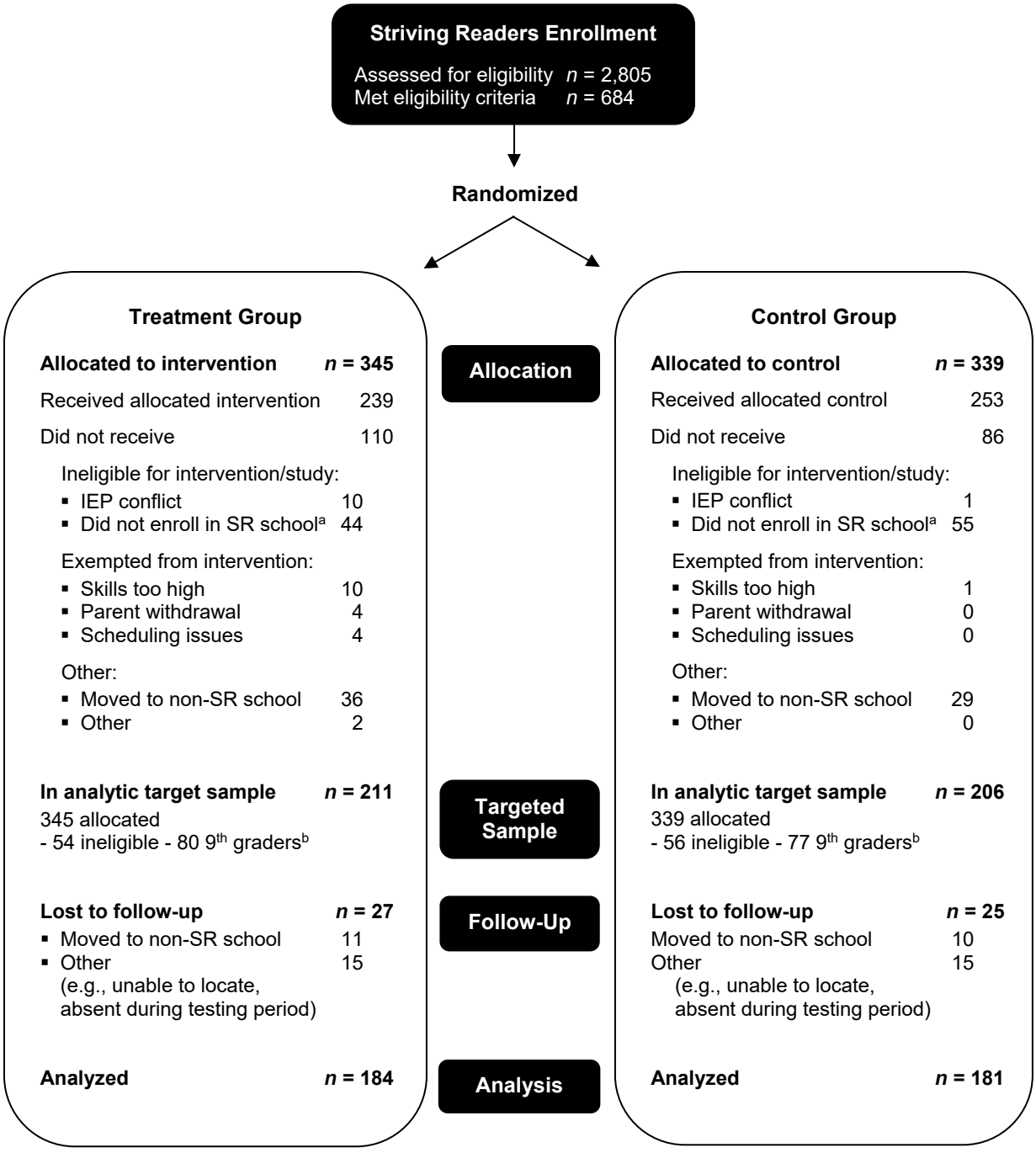

Note. IEP = individualized education program. SR = Striving Readers.
[a]These students never learned their group assignment. [b]Crossovers = 9. [c]Students in Grade 9 do not take the OAKS and were removed from the analytic sample.

**Exhibit 14**
**Targeted Intervention Random Assignment: Year 2 OAKS Sample**



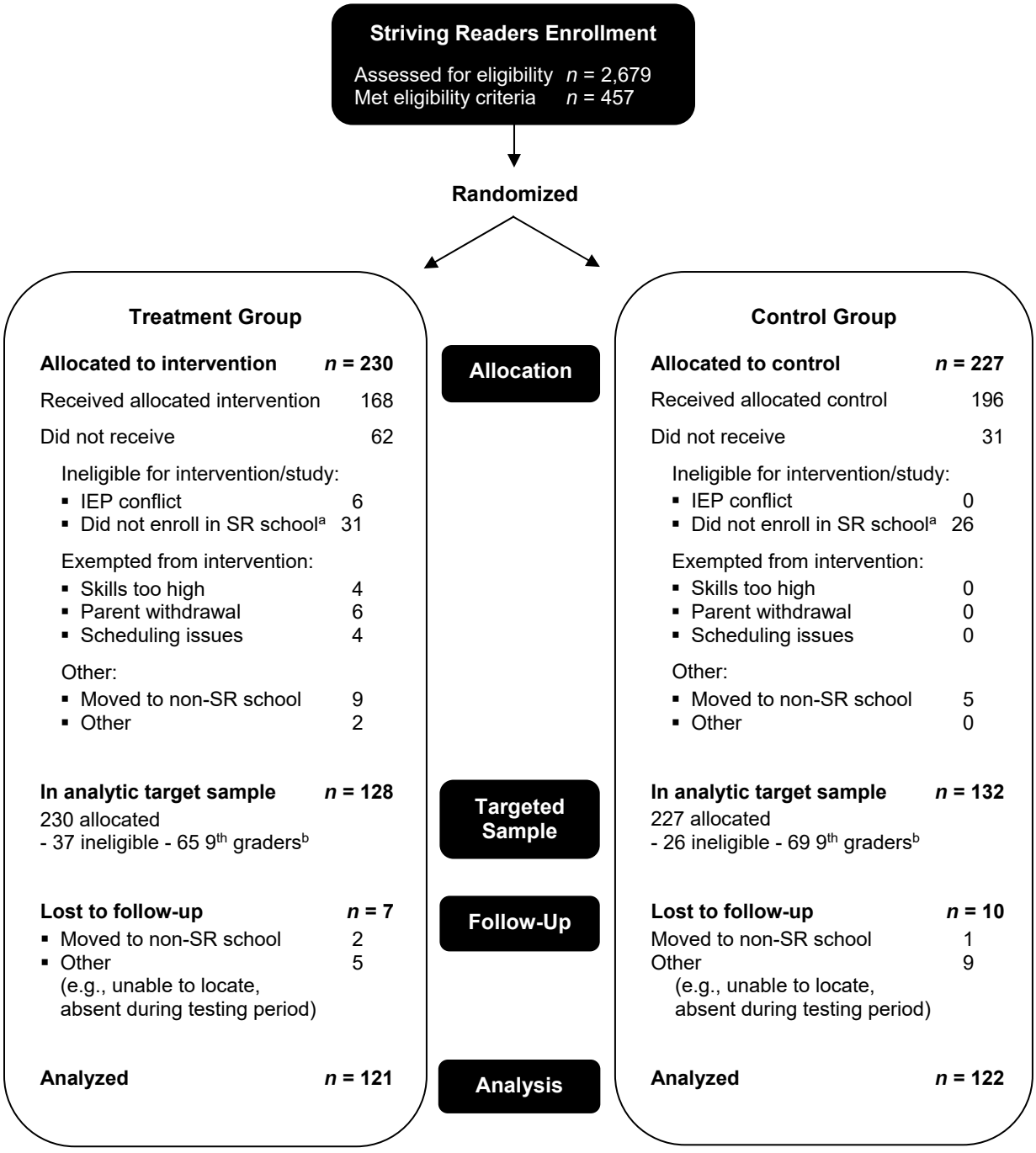

*Note.* IEP = individualized education program, SR = Striving Readers.
[a]These students never learned their group assignment. [b]Students in Grade 9 do not take the OAKS and were removed from the analytic sample.

**Exhibit 15**
**Targeted Intervention Random Assignment: Year 3 OAKS Sample**




*Note.* IEP = individualized education program, SR = Striving Readers.
[a]These students never learned their group assignment. [b]Students in Grade 9 do not take the OAKS and were removed from the analytic sample.

**Exhibit 16**
**Targeted Intervention Random Assignment: Year 4 OAKS Sample**




*Note.* IEP = individualized education program, SR = Striving Readers.
[a]These students never learned their group assignment. [b]Students in Grade 9 do not take the OAKS and were removed from the analytic sample.

### *Description of the Counterfactual*

The counterfactual for Xtreme Reading varied across schools because students in the control group did not receive a uniform alternative to the Xtreme Reading intervention. The difference was greatest between the middle schools (which used a replacement model) and the high schools (which used an add-on model). In the middle schools Xtreme Reading was delivered during the language arts and social studies instruction block, and the primary difference between the treatment and control groups was lower student-teacher ratios in the classes taught by the Xtreme Reading teachers. Because Xtreme Reading was an elective course at the high school level, the control group students enrolled in a wide range of other electives across all content areas. In Year 1, 9 control group students were accidently placed in treatment classes, but this problem did not occur in subsequent years. Xtreme Reading teachers did not teach control group students, thus preventing exposure of the control group to the treatment.

### *Data Collection Plan*

The impact evaluation data analysis compared the treatment and control groups on 3 student outcomes:

- **GRADE test scores**—The GRADE is administered in the fall and spring to all treatment and control group students in Grades 7–10. The GRADE is a highly reliable test (For Total Reading on Level M, alpha [internal consistency] ranges from .94–.95; test-retest reliability ranges from .88–.92, and alternate forms reliability ranges from .82–.90, depending on grade level). Controlling for fall (pretest) scores, the data analysis examines NCE scores in the spring and the percentage of students reading at grade level in the spring. NCEs are normalized standard scores with a mean of 50 and a standard deviation of 21.06, and have the same range as percentiles (1–99).
- **OAKS scores**—The OAKS is a computer adaptive test administered up to 3 times per year in Grades 7, 8, and 10 to assess whether students have met state achievement standards in reading and other subjects. Controlling for scores from the previous testing period[5] (pretest), the analysis compares NCE scores for treatment and control group students. Standard reliability data for the OAKS are not available.

[5]The OAKS is not administered to Grade 9. Pretest data are based on scores from the previous testing period (i.e., Grade 6 for Grade 7 students, Grade 7 for Grade 8 students, and Grade 8 for Grade 10 students).

- **Motivation for reading**—A survey that measures motivation to read and confidence in reading skills is administered in fall and spring to all treatment and control group students. The evaluation team modified the measure between Years 1 and 2, eliminating unreliable items (i.e., reading work avoidance items and reading efficacy items) and adding new items (i.e., social items and extrinsic and intrinsic motivation items). In addition, the response options were expanded from 4 (1 = *very different from me*, 2 = *a little different from me*; 3 = *a little like me*; 4 = *a lot like me*) in Year 1 to 6 (1 = *strongly disagree* to 6 = *strongly agree*) in Years 2 through 4. An overall Motivation for Reading mean score was computed such that the Overall Motivation for Reading score ranged from 1 to 4 in Year 1 and from 1 to 6 in Years 2 through 4. In Year 1 the Overall scale contained 23 items, and Cronbach's $\alpha$ was .90; in Years 2 through 4 the scale contained 28 items and Cronbach's $\alpha$ was .94.

Student data for assessing the effectiveness of the targeted intervention were collected in accordance with the schedule in Exhibit 17.

**Exhibit 17**
**Targeted Intervention Impact Data Collection Schedule**

| Data Element or Instrument | Time Point |
|---|---|
| Student demographic data | Fall |
| GRADE | Fall and spring of treatment year |
| Follow-up GRADE—Xtreme Reading students Grades 8–11 | Spring following program exit |
| OAKS | Spring each year for Grades 7, 8, and 10 |
| Student survey—treatment and control groups | Fall and spring of treatment year |

### *Summary of Analytic Approach*

An intent-to-treat statistical model—a framework in which participants are analyzed within their initial random assignment group regardless of whether they actually received treatment—was used in this study. Because students are clustered within schools, a multilevel model was used to estimate the impact of the intervention on spring outcome NCE scores while controlling for baseline NCE, ethnicity, English language proficiency, grade level, and cohort. Two models are presented in this section: (a) a model for the GRADE analytic sample and (b) a model for the OAKS analytic sample. Each model was run for the overall sample and separately for the

middle school and high school samples. Additionally, the GRADE analytic sample model was run with GRADE as the outcome and with OAKS as the outcome.

Portland Public Schools provided individual student demographic data for grade level, sex, ethnicity, special education status, and English language proficiency. Using backwards elimination, covariates that uniquely contributed to variance in outcome (with $p < 0.20$) were retained. Demographic covariates retained included ethnicity (African American, Hispanic), English language proficiency, and grade level. No data were missing for any of the demographic covariates.

### GRADE Analytic Sample

Approximately 11% of the students in the sample were missing GRADE pretest scores. Because missing data were not randomly distributed across schools and grades, a dummy variable approach was used to address the missing pretest scores, thereby allowing an examination of differences in posttest scores between students with pretest data and those without. The 2-level model for estimating the impact of the intervention on change in GRADE outcomes is specified below.

Level 1 Model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(Trt_{ij}) + \beta_{2j}(Pre_{ij}) + \beta_{3j}(D_{ij}) + \beta_{4j}(Black_{ij}) + \beta_{5j}(Hispanic_{ij}) + \beta_{6j}(ELL_{ij}) + \beta_{7j}(Cohort_{ij}) + \sum \beta_{8j}(Grade_{ij}) + \varepsilon_{ij}$$

Level 2 Model:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$
$$\beta_{4j} = \gamma_{40}$$
$$\beta_{5j} = \gamma_{50}$$
$$\beta_{6j} = \gamma_{60}$$
$$\beta_{7j} = \gamma_{70}$$
$$\beta_{8j} = \gamma_{80}$$

where:

$\varepsilon_{ij}$ = the random error term for student *i* in school *j*

$\mu_{0j}$ = the random intercept term for school *j*

and:

$Y_{ij}$ = the follow-up GRADE NCE score for student *i* in school *j*

$Trt_{ij}$ = the treatment indicator for student *i* in school *j* (0 = control, 1 = treatment)

$Pre_{ij}$ = the pretest GRADE NCE score for student *i* in school *j* (coded as 0 if pretest GRADE score is missing)

$D_{ij}$ = the indicator for missing pretest data for student *i* in school *j* (0 = non-missing, 1 = missing)

$Hispanic_{ij}$ = the ethnicity indicator for student *i* in school *j* (0 = non-Hispanic, 1 = Hispanic)

$Black_{ij}$ = the ethnicity indicator for student *i* in school *j* (0 = non-Black, 1 = Black)

$ELL_{ij}$ = the English language proficiency indicator for student *i* in school *j* (0 = proficient, 1 = not proficient)

$Cohort_{ij}$ = the set of dummy coded cohort indicators for student *i* in school *j*

$Grade_{ij}$ = the set of dummy coded grade level indicators for student *i* in school *j*

**OAKS Analytic Sample**

Approximately 11% of the students in the sample were missing OAKS pretest scores. Missing data were not randomly distributed across schools and grades; thus a dummy variable approach was used to address the missing pretest scores, allowing an examination of differences in posttest scores between students with pretest data and those without. The 2-level model for estimating the impact of the intervention on change in OAKS outcomes is specified below.

Level 1 Model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(Trt_{ij}) + \beta_{2j}(Pre_{OAKSij}) + \beta_{3j}(D_{OAKSij}) + \beta_{4j}(Pre_{GRADEij}) + \beta_{5j}(D_{GRADEij}) + \beta_{6j}(Black_{ij}) + \beta_{7j}(Hispanic_{ij}) + \beta_{8j}(ELL_{ij}) + \beta_{9j}(Cohort_{ij}) + \sum \beta_{10j}(Grade_{ij}) + \varepsilon_{ij}$$

Level 2 Model:

$\beta_{0j} = \gamma_{00} + \mu_{0j}$

$\beta_{1j} = \gamma_{10}$

$\beta_{2j} = \gamma_{20}$

$\beta_{3j} = \gamma_{30}$

$\beta_{4j} = \gamma_{40}$

$\beta_{5j} = \gamma_{50}$

$\beta_{6j} = \gamma_{60}$

$\beta_{7j} = \gamma_{70}$

$\beta_{8j} = \gamma_{80}$

$\beta_{9j} = \gamma_{90}$

$\beta_{10j} = \gamma_{100}$

where:

$\varepsilon_{ij}$ = the random error term for student *i* in school *j*

$\mu_{0j}$ = the random intercept term for school *j*

and:

$Y_{ij}$ = the follow-up OAKS NCE score for student *i* in school *j*

$Trt_{ij}$ = the treatment indicator for student *i* in school *j* (0 = control, 1 = treatment)

$Pre_{OAKSij}$ = the baseline OAKS NCE score for student *i* in school *j* (coded as 0 if baseline OAKS score is missing)

$D_{OAKSij}$ = the indicator for missing baseline OAKS data for student *i* in school *j* (0 = non-missing, 1 = missing)

$Pre_{GRADEij}$ = the pretest GRADE NCE score for student *i* in school *j* (coded as 0 if pretest GRADE score is missing)

$D_{GRADEij}$ = the indicator for missing pretest GRADE data for student *i* in school *j* (0 = non-missing, 1 = missing)

$Hispanic_{ij}$ = the ethnicity indicator for student *i* in school *j* (0 = non-Hispanic, 1 = Hispanic)

$Black_{ij}$ = the ethnicity indicator for student *i* in school *j* (0 = non-Black, 1 = Black)

$ELL_{ij}$ = the English language proficiency indicator for student *i* in school *j* (0 = proficient, 1 = not proficient)

$Cohort_{ij}$ = the set of dummy coded cohort indicators for student *i* in school *j*

$Grade_{ij}$ = the set of dummy coded grade level indicators for student *i* in school *j*

## Description of the Combined Samples

### *GRADE Analytic Sample*

Exhibit 18 summarizes the characteristics of the treatment and control group students in the GRADE analytic sample. Overall, 52% of the Xtreme Reading participants were male, 75% were non-White, 25% received special education services, and 28% were English language learners. Overall, 53% of the control group participants were male, 76% were non-White, 26% received special education services, and 25% were English language learners. The evaluation team conducted equivalence tests on key factors to determine whether differences between the GRADE analytic sample treatment and control groups existed at baseline. Overall, the groups were equivalent in terms of sex, ethnicity, special education status, English language proficiency, and baseline OAKS scores, but students in the treatment group scored significantly higher than the students in the control group on the fall GRADE (*M*s = 28.40 and 26.70 for treatment and control groups, respectively). When broken out by school level, the treatment and control groups were equivalent on all characteristics at each school level. See Appendix D for baseline equivalence tables broken out by school level.

**Exhibit 18**
**Baseline Equivalence of Treatment and Control Groups on Demographic Variables**

| | Percentage of Students at Each School Level | | | | | |
|---|---|---|---|---|---|---|
| | Treatment | | | Control | | |
| **Characteristic** | **MS** | **HS** | **Total** | **MS** | **HS** | **Total** |
| **Sex** | | | | | | |
| Male | 53 | 52 | 52 | 53 | 52 | 53 |
| Female | 47 | 48 | 48 | 47 | 48 | 47 |
| **Ethnicity** | | | | | | |
| White | 25 | 24 | 25 | 24 | 24 | 24 |
| American Indian | 2 | 2 | 2 | 3 | 2 | 2 |
| Hispanic | 31 | 26 | 29 | 36 | 25 | 31 |
| African American | 23 | 34 | 28 | 23 | 36 | 29 |
| Asian | 18 | 11 | 15 | 13 | 11 | 12 |
| **Special education services** | 28 | 21 | 25 | 26 | 27 | 26 |
| **ELL services** | 36 | 20 | 28 | 32 | 19 | 25 |

*Note.* Percentages are for GRADE analytic sample. Treatment total *n* = 756; Middle school = 401; High school = 355. Control total *n* = 823; Middle school = 421; High school = 402. ELL = English language learner. A significant difference in percent Hispanic was found at the middle school level.

### *OAKS Analytic Sample*

The evaluation team conducted equivalence tests on key factors to determine whether differences between the OAKS analytic sample treatment and control groups existed at baseline. For the overall sample, the groups were equivalent in terms of sex, English language proficiency, and baseline GRADE and OAKS scores. However, the treatment group students in the overall sample were more likely than the control group students to be Hispanic. At the middle school level, the treatment group students performed better than the control group students on fall GRADE (*M*s = 27.03 and 24.84 for the treatment and control group, respectively), but there were no differences between treatment and control group students at the high school level. Baseline equivalence results for the OAKS analytic sample are broken out by school level in Appendix D.

## Impact on Students

### *GRADE Analytic Sample*

For the GRADE analytic sample, the impact of Xtreme Reading was evaluated in terms of spring GRADE NCE (controlling for fall GRADE NCE), OAKS NCE (controlling for baseline OAKS NCE), the percentage of students reading at grade level on the GRADE in the spring, and spring student reading motivation (controlling for fall reading motivation).

#### GRADE NCE Outcomes

A multilevel model was used to estimate the impact of the intervention on spring GRADE NCE while controlling for fall GRADE NCE, grade level (Grades 7–10), ethnicity (African American/non-African American, Hispanic/non-Hispanic), and English language proficiency (proficient/not proficient). Additionally, the analysis controlled for presence of fall GRADE NCE score (present/missing) to control for systematic differences in students' completion of the pretest. The data analysis revealed a significant intervention effect for the overall sample: the treatment group students had significantly higher outcomes than the control group students, $\beta = 3.34$, $p < .001$. A significant treatment effect was also present for both the middle school sample, $\beta = 4.39$, $p < .001$ and the high school sample, $\beta = 2.07$, $p < .05$. Exhibit 19 presents the impact estimates for the overall sample and by school level.

**Exhibit 19**
**GRADE Total Reading NCE Impact Estimates**

| | | Estimated Impact | | | |
|---|---|---|---|---|---|
| **Group** | ***n*** | **Impact (β)** | ***SE*** | **Effect Size** | ***p*** |
| Middle school | 822 | 4.39 | 0.69 | .29 | .000 |
| High school | 757 | 2.07 | 0.98 | .12 | .035 |
| Overall | 1,579 | 3.34 | 0.56 | .21 | .000 |

Using Glass's Δ method (with the impact estimate as the numerator and the control group standard deviation as the denominator) to calculate the standardized effect size, the standardized effect size for the overall sample was .21; at the middle school level the standardized effect size was .29, and at the high school level the standardized effect size was .12. Exhibit 20 displays GRADE descriptive statistics (means, standard deviations, and range) in NCEs for the overall sample and by school level.

**Exhibit 20**
**GRADE Total Reading Descriptive Statistics in NCEs**

| | | Middle School | | | High School | | | Overall | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Test** | **Group** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** |
| Pretest | Treatment | 26.94 | 13.54 | 1–75 | 30.20 | 16.10 | 1–74 | 28.40 | 14.82 | 1–75 |
| | Control | 24.95 | 15.11 | 1–65 | 28.73 | 15.76 | 1–78 | 26.70 | 15.52 | 1–78 |
| Posttest | Treatment | 33.36 | 14.14 | 1–81 | 31.89 | 18.79 | 1–85 | 32.67 | 16.49 | 1–85 |
| | Control | 27.44 | 15.07 | 1–68 | 28.95 | 17.20 | 1–81 | 28.18 | 16.15 | 1–81 |

*Note.* Posttest treatment total *n* = 756; Middle school total *n* = 401; High school total *n* = 355. Posttest control total *n* = 823; Middle school total *n* = 421; High school total *n* = 402.

To determine whether the intervention differentially affected students' development of vocabulary and comprehension skills, separate impact analyses were conducted with GRADE Vocabulary and Comprehension subtest NCE scores as outcomes. The results of both analyses were similar to those seen for Total Reading NCE scores. Impact estimates for these subtests are reported in Appendix G.

To assess whether students who were missing pretest scores differed from students who were not missing pretest scores, the impact model controlled for presence of fall GRADE NCE score (present/missing). Results showed no differences in outcome scores between students who

were and students who were not missing pretest scores. This finding was true for the overall sample, the middle school sample, and the high school sample.

### Reading at Grade Level

A total of 454 treatment group students and 492 control group students were reading 2 or more years below grade level on the GRADE at the time of the pretest. Of this group, a significantly greater percentage of treatment than control group students (45% and 35%, respectively) improved by one grade level or more after receiving the intervention, $\chi^2$ (1, $n$ = 944) = 9.80, $p < .01$. These differences were moderated by school level. Although no significant differences between treatment and control group students were evident at the high school level (35% and 38%, respectively), a significantly greater percentage of treatment than control group students at the middle school level (52% and 32%, respectively) improved by one grade level or more after receiving the intervention, $\chi^2$ (1, $n$ = 535) = 21.56, $p < .001$.

Only 4% of the 454 treatment group students and 2% of the 492 control group students who were reading 2 or more years below grade level at the pretest were reading at or above grade level[6] at posttest. Overall and at the high school level there were no significant differences between the treatment and control groups, but at the middle school level the percentage of treatment group students reading at grade level at posttest was significantly higher (4% and 1%, respectively). Of all 756 treatment and 823 control group students in the GRADE analytic sample, only 92 treatment and 56 control group students (12% and 7%, respectively) were reading at or above grade level at posttest.

### Motivation for Reading

A multilevel model similar to that used to estimate the impact of the intervention on spring GRADE NCE was used to estimate the impact of the intervention on spring motivation for reading scores. The data analysis revealed no significant differences between the treatment and control groups in terms of change in overall reading motivation for Cohort 1. A significant intervention effect was evident, however, for the overall Years 2 through 4 combined sample: the treatment group students had significantly higher outcomes than the control group students, $\beta = 0.12$, $p < .05$. Exhibit 21 presents the impact estimates for the overall sample and by school level for Years 2 through 4.

[6]Calculation of posttest grade equivalence used students' grade level for the coming fall.

**Exhibit 21**
**Motivation for Reading NCE Impact Estimates for Years 2 through 4**

| | Estimated Impact | | | | |
|---|---|---|---|---|---|
| **Group** | ***n*** | **Impact (β)** | ***SE*** | **Effect Size** | ***p*** |
| Middle school | 418 | 0.13 | 0.09 | .14 | .136 |
| High school | 350 | 0.13 | 0.09 | .15 | .149 |
| Overall | 768 | 0.12 | 0.06 | .13 | .044 |

Because the motivation for reading measure was modified between Years 1 and 2, descriptive statistics are reported separately for Cohort 1 and Cohorts 2–4 combined. For Cohort 1 the motivation for reading in the fall (pretest) was the most significant predictor of motivation for reading in the spring (posttest), and virtually no change in motivation occurred for either group (treatment and control group *M*s = 2.73 and 2.71 for fall and spring, respectively). In contrast, in the Cohorts 2–4 combined sample, the treatment group had significantly greater gains in reading motivation than the control group, $\beta = 0.12$, $p < .05$ (treatment group *M*s = 3.44 and 3.54 for fall and spring, respectively; control group *M*s = 3.47 and 3.44 for fall and spring, respectively). The standardized effect size was .12. Exhibit 22 displays the means and standard deviations for each sample of students.

**Exhibit 22**
**Motivation for Reading Descriptive Statistics**

| | | Treatment | | | Control | | |
|---|---|---|---|---|---|---|---|
| **Sample** | **Term** | ***N*** | ***M*** | ***SD*** | ***N*** | ***M*** | ***SD*** |
| Cohort 1 | Fall | 203 | 2.73 | 0.53 | 224 | 2.73 | 0.55 |
| | Spring | 206 | 2.71 | 0.54 | 251 | 2.71 | 0.52 |
| Cohorts 2–4 | Fall | 277 | 3.44 | 0.93 | 387 | 3.47 | 0.97 |
| | Spring | 342 | 3.54 | 0.98 | 426 | 3.44 | 0.99 |

*Note.* Cohort 1 response values range from 1 = *very different from me* to 4 = *a lot like me*; Cohorts 2–4 response options range from 1 = *strongly disagree* to 6 = *strongly agree*.

### *OAKS Analytic Sample*

The impact of Xtreme Reading was evaluated for the OAKS analytic sample in terms of spring OAKS NCE (controlling for baseline OAKS NCE and fall GRADE NCE).

#### OAKS NCE Outcomes

A multilevel model was used to estimate the impact of the intervention on spring OAKS NCE while controlling for baseline OAKS NCE, fall GRADE NCE, grade level (Grades 7–10), ethnicity (African American/non-African American, Hispanic/non-Hispanic), and English language proficiency (proficient/not proficient). Additionally, the analysis controlled for presence of baseline OAKS NCE score (present/missing) and fall GRADE NCE score (present/missing) to control for systematic differences in students' completion of the pretest. The data analysis revealed a significant intervention effect for the overall sample, $\beta = 1.26$, $p < .05$. Likewise, there was a significant treatment effect for the middle school sample, $\beta = -1.69$, $p < .05$; however, a significant treatment effect was not present for the high school sample, $\beta = 0.29$, $p = .784$. Exhibit 23 presents the impact estimates for the overall sample and by school level.

**Exhibit 23**
**OAKS NCE Impact Estimates (OAKS Analytic Sample)**

| | Estimated Impact | | | | |
|---|---|---|---|---|---|
| **Group** | ***n*** | **Impact (β)** | ***SE*** | **Effect Size** | ***p*** |
| Middle school | 954 | 1.69 | 0.69 | .12 | .014 |
| High school | 514 | .29 | 1.06 | .02 | .784 |
| Overall | 1,468 | 1.26 | 0.58 | .09 | .031 |

Using Glass's Δ method (with the impact estimate as the numerator) to calculate the standardized effect size, the standardized effect size for the overall sample was .09; at the middle school level the standardized effect size was .12, and at the high school level the standardized effect size was .02. Exhibit 24 displays OAKS descriptives (means, standard deviations, and ranges) in NCEs for the overall sample and by school level.

**Exhibit 24**
**OAKS Descriptive Statistics in NCEs (OAKS Analytic Sample)**

| | | Middle school | | | High school | | | Overall | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Testing** | **Group** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** |
| Pretest | Treatment | 27.35 | 11.50 | 1–67 | 32.89 | 12.70 | 1–73 | 29.04 | 12.14 | 1–73 |
| | Control | 26.78 | 12.10 | 1–81 | 33.21 | 14.92 | 1–81 | 28.73 | 13.34 | 1–81 |
| Posttest | Treatment | 32.40 | 12.55 | 1–77 | 35.19 | 13.86 | 1–87 | 33.39 | 13.09 | 1–87 |
| | Control | 29.98 | 13.76 | 1–72 | 34.85 | 13.60 | 1–71 | 31.66 | 13.89 | 1–72 |

*Note.* Posttest treatment total $n$ = 732; Middle school total $n$ = 472; High school total $n$ = 260. Posttest control total $n$ = 736; Middle school total $n$ = 482; High school total $n$ = 254.

To assess whether students missing pretest scores differed from students not missing pretest scores in terms of the OAKS impact, the model controlled for presence of fall GRADE NCE score (present/missing) and pretest OAKS NCE score (present/missing). No differences in outcome scores were evident between students who were missing pretest GRADE NCE scores or OAKS NCE scores and students who were not for the overall, middle school, or high school samples.

## Implementation Effects on the Impact of the Intervention

The impact analysis of the intervention on spring GRADE NCE also revealed significant school level variability in treatment effects in both the overall sample and the middle school sample. To further explore the possible effects of implementation on intervention impact, the evaluation team conducted a nonexperimental analysis to address the following question: To what extent do teacher variables explain school level variability in treatment effects?

A multilevel model identical to that used to measure impact—but with 3 teacher level variables (aggregated to the school level) added into Level 2 of the model—was used to estimate the effects of implementation on the impact of the intervention on spring GRADE NCE. The 3 teacher level variables included in the model were fidelity of implementation (measured by the evaluation team through classroom visits), amount of professional development attended (collected from attendance records), and years of teaching experience (collected from teacher surveys). The 2-level model for estimating the impact of the intervention on change in GRADE outcomes is specified below.

Level 1 Model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(Trt_{ij}) + \beta_{2j}(Pre_{ij}) + \beta_{3j}(D_{ij}) + \beta_{4j}(Black_{ij}) + \beta_{5j}(Hispanic_{ij}) + \beta_{6j}(ELL_{ij}) + \sum \beta_{7i}(Grade_{ij}) + \varepsilon_{ij}$$

Level 2 Model:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(PDAtt) + \gamma_{12}(Fidelity) + \gamma_{13}(YrsTch) + \mu_{1j}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

$$\beta_{4j} = \gamma_{40}$$

$$\beta_{5j} = \gamma_{50}$$

$$\beta_{6j} = \gamma_{60}$$

$$\beta_{7j} = \gamma_{70}$$

Due to changes in teacher staffing from one year to the next, separate models were run for each implementation year. Significant variability in treatment effect was evident for Years 1 and 3, but not for Years 2 and 4. Thus the nonexperimental analysis was not conducted for the Years 2 and 4 samples. Addition of the 3 teacher variables in the Year 1 model revealed that teacher fidelity of implementation significantly contributed to the between-school variability in treatment effect, $\beta = 0.21$, $p < .05$, but teaching experience or amount of professional development made no significant contribution. Addition of the teacher variables in the Year 3 model revealed that none of the 3 variables—teacher fidelity of implementation, amount of professional development, or number of years of teaching experience—accounted for the between-school variability in treatment effect.

## Discussion

For the GRADE and OAKS analytic samples, multilevel analyses consistently revealed a significant intervention effect at the middle school level whereby the treatment group performed significantly better than the control group. The standardized effect size for the GRADE analytic sample at the middle school level was .29 for GRADE as the outcome and .15 for OAKS as the outcome. For the OAKS analytic sample with OAKS as the outcome, the effect size at the middle school level was .12. The impact of the intervention at the high school level was

significant only for the GRADE analytic sample with GRADE as the outcome (effect size = 0.12). For the high school GRADE analytic sample with OAKS as the outcome, the standardized effect size was .02, and for the OAKS analytic sample with OAKS as the outcome, the effect size was also .02. The differences in impact by school level are not surprising given the differences in program implementation: the middle schools had higher levels of teacher participation in the professional development, more Xtreme Reading teachers with a reading endorsement, higher fidelity of Xtreme Reading implementation in the classroom, and higher levels of teacher buy-in.

Supplemental analyses that examined the effects of the intervention on GRADE Vocabulary and Comprehension subtest NCE scores revealed results that were consistent with program effects on Total NCE score. This finding indicates that the intervention had similar effects on vocabulary and comprehension skill development. In terms of the other outcomes, there was a significant difference between the treatment and control groups in terms of percentage of students reading at grade level at posttest. Of the overall sample of students reading 2 or more years below grade level at the time of the pretest, a significantly greater percentage of treatment than control group students improved by one grade level or more after the receiving the intervention. This effect was observed for the middle school sample but not for the high school sample. In terms of reading motivation, no differences were evident between the treatment and control groups in change in reading motivation for Cohort 1. In the Cohorts 2–4 combined sample the treatment group did, however, show significantly greater gains in reading motivation than the control group but only for the overall sample; this was not evident for either the middle school or high school samples.

Because the intervention was implemented differently across the 10 Striving Readers schools, a nonexperimental analysis was conducted to examine between-school variability in treatment effects. These analyses revealed that for Years 1 and 3 but not Years 2 and 4, between-school variability in treatment effect was significant. Additional analyses that explored the possible effect of implementation on the impact of the intervention revealed that teacher fidelity of implementation, but not number of years of teacher experience or amount of professional development attended, significantly contributed to the between-school variability in treatment effects. This finding was true for the Year 1 sample, but in the Year 3 sample none of the teacher variables accounted for the between-school variability in treatment effect.

# Implementation of the Whole School Intervention

To provide context for interpreting the impact of the whole school intervention, the evaluation team assessed the implementation of the content enhancement routines in Years 1 through 5. Program staff expected implementation to improve as more teachers gained experience using the instructional strategies across content areas. Although various factors contributed to the fidelity of implementation, barriers to implementation also emerged. This section of the report describes the implementation evaluation design, the findings from Years 1 through 5, and the implications of these findings for the impact analyses.

## Implementation Evaluation Design

The evaluation team collected data for the whole school intervention implementation evaluation through surveys or interviews with the professional developers, the Striving Readers district leadership team, school administrators, school literacy coaches, and classroom teachers and by conducting classroom observations. Language arts and social studies teachers participated in the whole school intervention in Year 1; and they were joined by their math and science colleagues in Year 2; and arts, physical education, and health teachers in Year 3. The data were analyzed in terms of professional development inputs, classroom implementation fidelity, and teacher buy-in.

- **Professional development inputs** were defined for each school as the teachers' average group professional development attendance, participation in coaching with the professional developers, and participation in coaching with the school literacy coaches, and the qualifications of the school literacy coaches. The professional development offered in Year 1 included a summer training session, fall make-up training sessions, and periodic group workshops conducted during the school year. The professional development offered in Year 2 included a summer training session, fall make-up training sessions, and optional training in additional content enhancement routines. The professional development offered in Year 3 included a summer training session at each of the 2 new middle schools and fall make-up training sessions. In Year 3 ongoing professional development changed from being centralized to being school based so that individual school needs and interests would be better addressed. In Year 4, all professional development for content teachers was provided by district staff except for

the 2 middle schools that were in their second year of implementation. These 2 schools continued to receive professional development from KU staff. For the remaining schools, professional development was offered in the summer at a central location for any new or experienced content teachers who wished to participate. In Year 5, professional development was provided by district staff in the 6 middle schools. Most of this professional development was in the form of facilitated work time to assist content teachers in integrating the content enhancement routines in their lessons.

- **Classroom implementation fidelity** was defined as the average of the fidelity ratings that were part of the classroom observations conducted by the evaluation team in winter and spring in Year 1 and in fall and spring in Years 2 through 4; the percentage of required content enhancement routines completed (averaged across teachers at a school),[7] and teachers' self-reported frequency of content enhancement routine use in the month prior to the surveys administered in fall, winter, and spring (averaged across teachers at a school). Classroom observations were not conducted in Year 5.
- **Teacher buy-in** was evaluated by averaging teachers' survey responses in 2 domains: (a) their perceptions of the group professional development sessions and in-school coaching provided by the professional developers and school literacy coaches, which was measured by their agreement on a 5-point scale with items such as "The professional development on the content enhancement routines prepared me to use these routines effectively in my classroom" and "My school's Striving Reader's literacy coach has helped me to implement the content enhancement routines" and (b) their perceptions of the effectiveness of specific content enhancement routines as measured by their ratings on a 5-point scale.

## Implementation Evaluation Findings

Exhibit 25 summarizes the levels of implementation of the content enhancement routines at each of the participating schools in Years 1 through 5. The data in each category (professional development inputs, classroom implementation fidelity, and teacher buy-in) were averaged across teachers and content areas.

[7]Completion rates determined from school literacy coaches' records. In Year 1 a specific number of certain content enhancement routines was required. In Years 2 through 4 teachers could select the content enhancement routines, but completion of a total number was required.

## Exhibit 25
## Implementation of the Whole School Intervention

| | Professional Development Inputs | | | | | Classroom Implementation[a] | | | | Teacher Buy-In[b] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **School** | **Yr 1** | **Yr 2** | **Yr 3** | **Yr 4** | **Yr 5** | **Yr 1** | **Yr 2** | **Yr 3** | **Yr 4** | **Yr 1** | **Yr 2** | **Yr 3** | **Yr 4** | **Yr 5** |
| **Middle** | | | | | | | | | | | | | | |
| Alpha | M | M | H | M | H | M | M | M | L | L | M | H | H | H |
| Beta | M | M | M | L | H | M | M | M | M | M | M | M | M | H |
| Gamma | M | M | L | L | M | L | M | M | M | M | M | H | M | M |
| Delta | L | M | M | H | H | L | L | M | M | L | M | M | M | H |
| Epsilon | M | L | – | – | – | M[c] | M | – | – | M[d] | L | – | – | – |
| Omega | – | – | M | L | M | – | – | M | L | – | – | M | M | M |
| Zeta | – | – | L | M | H | – | – | M | M | – | – | M | H | H |
| **High** | | | | | | | | | | | | | | |
| Kappa | M | L | L | M | – | L[c] | L[c] | L | M | L[d] | L[d] | M | H | – |
| Lambda | L | M | M | L | – | L[c] | L[c] | L | L | L[d] | L[d] | M | M | – |
| Sigma | M | M | L | L | – | L[c] | L[c] | L | L | L[d] | L | M | L | – |
| Theta | M | L | L | L | – | L[c] | L[c] | L | vL[c] | L[d] | L | L | L | – |

*Note*. H = *high*, M = *medium*, L = *low,* vL = *very low.* Dashes indicate that school did not participate. See Appendix B for additional information on calculation of ratings.

[a]Classroom observations contributed 50% to classroom implementation ratings, and teacher participation varied across time points (e.g., winter, spring) and schools (i.e., range of 9% to 100% of teachers participating). In general, participation was higher and more consistent across time points in the middle schools, and lower and more variable in the high schools. [b]Teacher buy-in derived from teacher survey responses; some schools had lower teacher response rates than others. [c]School had low observation participation rates, which might limit their interpretation. [d]School had low survey response rate, which might limit their interpretation.

### *Professional Development*

In Year 1 40% of the middle school language arts or social studies teachers and 50% of the high school language arts teachers participated in an adequate number of professional development sessions. (Adequate participation was defined as participating in 76–100% of the professional development sessions offered.) None of the high school social studies teachers participated in adequate professional development. In Year 2 80% of the middle school teachers participated in adequate professional development. At the high school level 93% of the math teachers and 75% of the science teachers participated in adequate professional development. In contrast, only 50% of the high school language arts teachers and 59% of the high school social studies teachers participated in adequate professional development in Year 2. In other words, among the high schools math and science teachers generally participated in more professional development than their language arts and social studies colleagues. Participation at the middle school level increased significantly in Year 2. In Year 3 50% of the middle school teachers participated in an adequate number of professional development sessions. At the high school level, 75% of math and science teachers, 50% of language arts teachers, 25% of social studies teachers, and none of the arts, health, and physical education teachers participated in an adequate number of professional development sessions. In Year 4 all of the middle school teachers participated in an adequate number of professional development sessions. At the high school level 75% of math teachers, 50% of science and language arts teachers, and 25% of social studies and PE/health teachers participated in an adequate number of professional development sessions. In Year 5, school literacy coaches were dropped from the 4 high schools but continued at the 6 middle schools. Two-thirds of the middle school teachers participated in an adequate number of professional development sessions.

Because teachers were encouraged but not required to participate in coaching with the professional developers and the school literacy coaches, the amount of coaching the teachers received varied considerably each year. Overall, in Year 1 no schools met the target of 14 hours of contact per teacher with the professional developers (median of 2.8 hours) or the target of 14 hours of contact per teacher with the school literacy coaches (median of 11.2 hours). On average, the teachers who participated in coaching in Year 2 had more contact with the professional developers (median of 8.2 hours) and somewhat less contact with the school literacy coaches (median of 7.7 hours) compared to Year 1. Teachers who participated in coaching in Year 3 received a median of 6.8 hours of coaching from the professional developers

and 12.4 hours of coaching from the school literacy coaches. Coaching from the KU professional developers was not tracked in Year 4 because their contract ended midyear and most of their time onsite was spent working with the school Striving Readers literacy coaches. The school-level coaches continued to provide assistance to content teachers at the middle school level, but coaching at the high schools was infrequent. The amount of coaching received is a possible intervening variable for interpreting the fidelity of implementation of the whole school intervention. In terms of school literacy coach qualifications, in Year 1 60% of the middle schools and 100% of the high schools had literacy coaches with either a Master's degree or a reading endorsement or both. In Year 2 60% of the middle schools and 50% of the high schools had literacy coaches with a Master's degree or a reading endorsement or both, and in Year 3 these figures increased to 67% of the middle schools and 75% of the high schools respectively. In Year 4, 50% of the middle schools and 100% of the high schools had literacy coaches with a Master's degree or a reading endorsement or both, although 1 high school did not have a literacy coach at all. In Year 5, 50% of the middle school literacy coaches had a Master's degree or a reading endorsement or both.

#### *Classroom Implementation Fidelity*

Middle schools performed at similar levels in Years 1–4, and a majority displayed medium fidelity. At the high school level participation in classroom observations was very low in all years and caution should be exercised when interpreting these findings. Classroom implementation fidelity was, however, generally lower at the high school level than at the middle school level.

#### *Teacher Buy-In*

Teacher buy-in remained relatively stable at both school levels over the first 2 years but increased in Year 3. In Year 4, 60% of the schools maintained the same buy-in level, 20% improved, and 20% declined. In Year 5, 67% of the middle schools maintained the same buy-in level and 33% improved. The middle school teachers reported somewhat higher buy-in than the high school teachers in Years 1–4 (teacher buy-in was not assessed at the high schools in Year 5).

## Implications for the Impact Analyses

Interpretation of the impact of the whole school intervention is facilitated by an examination of the issues that emerged during the 5 years of Striving Readers implementation. This section

focuses on factors that facilitated the implementation of the content enhancement routines, differences in implementation across school levels, and barriers to implementation as planned.

### *Factors That Facilitated Implementation*

Although many teachers and school administrators were critical of the summer training in Year 1, most agreed that monthly site visits conducted by the professional developers were very beneficial. Principals reported productive interactions with the professional developers and received positive feedback from the teachers. Principals also appreciated the technology provided through the Striving Readers grant, the focus on low achieving students, the emphasis on planning and structuring content instruction, and the emphasis on teacher collaboration toward a common goal. Many teachers commented favorably on the content enhancement routines, reporting that they had helped them and their students better conceptualize and grasp content.

In Year 2 the summer training was reduced in duration from 5 days to 3 days, and teachers were grouped by content area, which reduced the training group sizes. Although teachers were not entirely satisfied with the professional developer's efforts to tailor the training to each content area, more content-appropriate examples were provided. Another improvement was the involvement of Teachers on Special Assignment in the selection of the content enhancement routines that were taught to the math and science teachers. School administrators believed that staff attitudes toward the Striving Readers program improved in Year 2, in part because the professional developers demonstrated responsiveness to the teachers' concerns and made improvements to the training and the support provided.

In Year 3 the summer training was provided to the third wave of content teachers: arts, physical education, and health. Teachers new to the Striving Readers schools also received training on the content enhancement routines in their subject areas. In addition, the district implemented a new model of support during Year 3. Instead of a team of external professional developers providing the professional development at each school, each school had a single professional developer or advocate assigned to their school. District staff considered this approach effective at some schools, but success depended to some extent on the involvement of school administrators.

In Year 4 summer training was provided to new and continuing content teachers who elected to attend. Consistent with the transition plans, the content enhancement routine training was

provided by PPS staff who had previously been certified as Strategic Instruction Model trainers. KU staff continued to provide ongoing support at the 2 middle schools that were added in Year 3 and provided limited support to the other Striving Readers schools during the first half of the school year. In Year 5, no summer training was provided in the content enhancement routines, but training and facilitated work time were offered throughout the school year. For the most part, only the middle school teachers participated.

The school year professional development offerings were also modified over time. In Year 1 the ongoing professional development was centralized, and teachers selected from a menu of training sessions on new content enhancement routines. In Year 2 the professional developers and school administrators selected ongoing professional development topics for individual schools. Some of these sessions were designated as facilitated work time, which allowed teachers to collaborate to develop content enhancement routines for common units. Both the professional developers and the school literacy coaches reported, however, that the teachers did not always use this time effectively. In terms of the ongoing professional development provided to the school literacy coaches, in Year 2 the professional developers spent more time supporting the school literacy coaches to improve their coaching skills. In addition, the district leadership team instituted twice-monthly meetings that provided additional opportunities for the school literacy coaches to receive professional development and share their successes and challenges. District staff continued monthly meetings with the school literacy coaches in Years 3 and 4, and provided two days of training for middle school literacy coaches in Year 5.

Another factor facilitating implementation of the whole school intervention in Year 2 was the fact that some teachers noticed progress in their students' ability to organize ideas and concepts, which resulted in improvements in their writing. Teachers reported that students appreciated the consistent use of content enhancement routines across content areas, which facilitated learning. In addition, in Year 2 the district leadership team and the professional developers planned and began implementing a series of activities that allowed district staff to assume more responsibility for the professional development for both the targeted and whole school interventions. Staff who completed the series received certification in the reading strategies employed in Xtreme Reading (6 participants) or the content enhancement routines (9 participants).

In Year 3 Portland Public Schools staff played a greater role in providing professional development as part of an effort to become certified Strategic Instruction Model trainers. District staff also worked more with school literacy coaches to resolve issues with school administrators.

Teachers reported that the more they used the content enhancement routines, the more organized and consistent they became, the more they were able to focus on important content, and the more benefits they observed among their students. Teachers appreciated the time they had to develop routines together, the variety of routines that they had been taught to use, the flexibility of the routines, and the technology (document cameras, projectors, computers) they had received to assist with lesson presentation. Teachers noted that the content enhancement routines were especially helpful for students who struggled with organization or language and helped students determine the relative importance of content.

In Years 4 and 5 teachers who used the content enhancement routines continued to see benefits for themselves and their students due to better organized instruction, equipment for delivering instruction, facilitated work time to work with other teachers to develop new units, and assistance from the school literacy coach. Some principals felt that teacher use of the content enhancement routines increased as they learned about new content enhancement routines and had more experience with using them. Principals especially liked the content enhancement routines that involved teacher planning and organizing of curriculum and lessons to meet state and district standards.

### *Variation in Implementation by School Level*

On average, in Year 1 high school social studies teachers participated in less group professional development than high school language arts teachers or middle school teachers. Teacher participation in coaching (provided by either the professional developers or the school literacy coaches) varied considerably, and no clear patterns were evident except that the middle school teachers were more likely to receive coaching from the professional developers. In Year 2 the middle school teachers were again more likely to receive coaching than the high school teachers. For example, no middle schools reported an average of fewer than 5 hours of teacher contact with their school literacy coaches, whereas at least 50% of the high schools reported an average of fewer than 5 hours of contact. A subset of high school math teachers did, however, participate in a large amount of coaching. The addition of 2 new middle schools in Year 3 further increased the disparity between the middle and high schools. The new middle schools had strong administrative support for Striving Readers and ensured that 100% of their teachers participated in classroom observations and coaching. Overall, in Year 3 none of the middle schools averaged fewer than 6 hours of teacher contact with the professional developers or fewer than 12 hours of teacher contact with the school literacy coaches. In contrast, the high

schools averaged fewer than 6 hours of teacher contact with both the professional developers and the school literacy coaches.

At the high school level the classroom implementation fidelity data for all years must be interpreted with caution due to low classroom observation participation rates (26% in winter and 19% in spring in Year 1, 24% in fall and 17% in spring in Year 2, and 21% in fall and 9% in spring in Year 3). The fidelity ratings across all schools for Years 1 and 2 were generally below 79%. In Year 3, however, the fidelity ratings at the middle school level improved to the point that 50% of the middle schools had a fidelity rating in the 80–89% range. In terms of teacher buy-in, the high school teachers reported lower buy-in than the middle school teachers in all years. The professional developers, school principals, and school literacy coaches suggested that the high school math and science teachers demonstrated more buy-in and greater implementation fidelity than their language arts and social studies colleagues.

Other whole school intervention implementation patterns were evident by school level. In general, the middle school teachers were more receptive to implementing the content enhancement routines, working with the professional developers and school literacy coaches, and allowing classroom observations. This receptivity occurred despite extenuating circumstances in 4 of the 5 middle schools in Year 1: one school was preparing to close, 2 schools were preparing to merge with elementary schools, and one school was preparing to become an all-girls school. The implications of such changes included higher than usual staff turnover (which resulted in the loss of school literacy coaches and teachers with experience using content enhancement routines in Year 2), administrator turnover, and school initiatives that competed for teachers' and administrators' attention.

### *Barriers to Implementation*

Data on implementation barriers were gathered through professional developer interviews conducted in fall, winter, and spring; school administrator interviews conducted in fall and spring; school literacy coach surveys and interviews conducted in fall, winter, and spring; and teacher surveys administered in fall, winter, and spring. In Year 1 school administrators believed that the primary barriers to implementation included their own lack of knowledge of the targeted and whole school interventions, the competing demands of multiple reforms and programs, the late hiring of key staff, some teachers' unwillingness to participate in the Striving Readers program, and conflicts between the professional development requirements for Striving Readers

and other initiatives. Principals believed that implementation would have been more successful if all teachers (rather than language arts and social studies teachers only) had been involved in the program in Year 1 and issues involving the teachers union had been resolved earlier. “The high school does not have much of a choice in staffing—it is directed by our union,” remarked one principal, “We don’t get to pick the teachers or literacy coaches we want—it’s all based on licensure and seniority.” The principals also regretted that the district’s summer training for principals was scheduled at the same time as the Striving Readers summer training (in which they were also expected to participate).

Teachers suggested that the Year 1 summer training could have covered all of the content in much less time and complained that the professional developers did not take into account the knowledge they already had. Teachers also cited insufficient time for collaboration and the absence of clearly communicated expectations as barriers to implementation. The professional developers agreed that the district’s failure to clearly communicate implementation expectations was a problem in Year 1, along with a lack of consequences for teachers who refused to participate. They suggested that more administrative support would have been helpful, particularly to recognize teachers who successfully implemented the program. The professional developers also posited that teachers in Portland Public Schools lacked planning experience and consequently struggled to incorporate content enhancement routine planning into their daily schedules. The lack of a district curriculum for most content areas also thwarted implementation. Finally, some teachers developed content enhancement routines only for lower level or simple concepts rather than the complex concepts for which the routines are most helpful.

In Year 2 the district leadership team, school administrators, and the professional developers all noted improved teacher attitudes toward the Striving Readers program although many of the same implementation barriers persisted. In addition, math and science teachers were dealing with new curricula, some schools were in corrective action for not making Adequate Yearly Progress, and most schools were attempting to implement multiple school or instructional improvement initiatives at the same time. The district leadership team was concerned about the lack of accountability for teachers and principals, the lack of coordination with other initiatives, and continuing problems with the professional developers. These problems included failure to model best practices during professional development sessions, lack of clear expectations for teachers in the facilitated work sessions held during the school year, poor planning and coordination of professional development across school sites, and insufficient clarity in

procedures for Portland Public Schools staff to become certified Strategic Instruction Model professional developers.

In Year 3 teacher attitudes toward the Striving Readers program continued to improve but issues remained with respect to the quality and consistency of the professional development provided by the external professional developers. The district leadership team suggested that some of the professional developers were more skilled than others. Furthermore, although training in embedded strategies had been part of Portland's Striving Readers professional development plan since the inception of the grant, the external professional developers did not develop relevant materials or a plan for training teachers to use embedded strategies in the classrooms. Due to staffing changes , the district leadership team also lacked a program manager for part of Year 3. Teachers repeatedly requested more group planning time to develop content enhancement routines for their content areas, additional examples of exemplary routines, and improvements to the GIST software for developing routines. Some teachers never understood how content enhancement routines were supposed to improve student reading comprehension.

By Year 4, most of the barriers to implementation that had emerged in prior years remained. The Striving Readers program continued to be less successful at schools with weak building leadership concerning instruction. Some literacy coaches were frustrated that their administrators did not require teachers to use content enhancement routines and did not themselves seem interested in learning more about the program. Literacy coaches at the high schools found it particularly difficult to get teachers to work with them. Some administrators continued to have problems with hiring the best staff for the literacy coach position because of union constraints and others noted that it was difficult to get teacher buy-in when there wasn't strong evidence that use of the content enhancement routines led to better student outcomes. Both teachers and administrators continued to feel that Striving Readers was just one more competing demand for their limited time.

In Year 5, literacy coaches were employed only in the 6 continuing middle schools. The 10 new middle schools that were added for the targeted intervention were not trained to participate in the whole school intervention, and literacy coaches at the high school level were discontinued because of low interest from high school teachers in using content enhancement routines.

# Impact of the Whole School Intervention

The evaluation team assessed the impact of the whole school intervention in terms of changes over time in student reading achievement across the Striving Reader school populations as measured by the statewide achievement test (OAKS). This section of the report describes the impact evaluation design and the characteristics of schools included in the evaluation.

## Study Design

The whole school intervention impact evaluation addresses one evaluation question:

- To what extent do students in Grades 6–10 improve their OAKS reading scores after implementation of the content enhancement routines?

This section describes the impact evaluation sampling plan, the data collection plan, and the analytic approach that was used to answer the impact evaluation question.

### *Sampling*

The sample for the whole school intervention includes 9 Striving Readers schools.[8] Trends in schools' average OAKS scores (averages of all students in Grades 6–10) were tracked over time, beginning 3 years prior to the Striving Readers grant through the end of the current evaluation period (spring 2009). If the evaluation team assumes that over 7 years 63 observations of 9 schools will be conducted, an effect size of .55 should be detectable with .80 power when alpha is set at .05, the intraclass correlation is .05, and a 2-tailed test is used.

### *Data Collection*

The data analysis uses one student outcome measure:

- **OAKS scores**—The OAKS is administered in the spring to all students in Grades 3, 4, 5, 6, 7, 8, and 10.

Exhibit 26 shows the data collection schedule.

[8]The whole school intervention was initially implemented in 9 schools. Two additional schools began Striving Readers implementation in Year 3 and had only 1 year of postintervention data at the time of this report. Two of the initial schools closed after the first year of Striving Readers with only 1 year of postintervention data; these 2 schools were excluded from the analysis.

**Exhibit 26**
**Whole School Intervention Impact Data Collection Schedule**

| Data Element or Instrument | Time Point |
|---|---|
| Student demographic data | Fall |
| OAKS—Grades 6, 7, 8, 10 | Spring |

### *Analytic Approach*

To assess the schoolwide effects of the whole school intervention, a longitudinal multilevel analysis was conducted with repeated observations collected on schools before and after implementation of the intervention. The approach is a baseline mean projection model that used 3 baseline years and 3 years of implementation (Time = 0,0,0,0,1,2,3, with Time centered at the 2006–2007 school year). A piecewise term, Difference (coded as 0 prior to the whole school intervention implementation and one after the onset of implementation) was used to estimate the change from preimplementation to postimplementation. School level variables were included at Level 2. Below is the model specification for the proposed multilevel analysis.

At Level 1 (within subjects) the model is:

$$Y_{ij} = \beta_{0i} + \beta_{1i}\,\text{Time}_{ij} + \beta_{2i}\,\text{Difference} + r_{ij}$$

Where

- $Y_{ij}$ = reading achievement for school *i* at time *j*
- $\beta_{0i}$ = the baseline mean reading achievement for school *i*
- $\beta_{1i}$ = the postimplementation rate of change in reading achievement for school *i*
- $\beta_{2i}$ = the "jump" in reading achievement between pre- and postimplementation periods for school *i*
- $r_{ij}$ = error for school *i* at time *j*

At Level 2 (between subjects) the model is:

$$\beta_{0i} = \gamma_{00} + u_{0i}$$
$$\beta_{1i} = \gamma_{10}$$
$$\beta_{2i} = \gamma_{20}$$
$$\beta_{3i} = \gamma_{30}$$

Where:

- $\gamma_{00}$ = the baseline mean reading achievement

$u_{0i}$ = the difference between the baseline mean and individual school reading achievement for school *i*

$\gamma_{10}$ = the mean postimplementation rate of change in school reading achievement

$\gamma_{20}$ = the mean jump in reading achievement between the pre- and postimplementation periods

Portland Public Schools provided the following school demographic data: percentage of students eligible for free- or reduced-price lunch, ethnicity percentages, percentage of students receiving special education services, and percentage of English language learners. Using backwards elimination, covariates that uniquely contributed to variance in outcome (with $p < 0.20$) were retained. Covariates retained included ethnicity (African American, Hispanic) and school level.

## Description of the Whole School Intervention Samples

This section describes the characteristics of the schools, teachers, classrooms, and students participating in the whole school intervention.

### *School Characteristics*

Enrollment varied considerably across the middle and high schools. Among the 6 middle schools, Grades 6–8 enrollment in Year 5 ranged from approximately 150 in the smallest schools to 250 in the largest. Among the 4 high schools, Grades 9–12 enrollment in Year 5 ranged from approximately 415 in the smallest school to just over 1,000 in the largest. The percentage of English language learners ranged from 5% to 42%, and the percentage of students eligible for free- or reduced-price lunch ranged from 40% to 91%. The proportion of non-White students ranged across schools from 36% to 87%. The percentage of students receiving special education services ranged from 14% to 24%. Exhibit 27 shows the characteristics of the 10 schools that implemented the whole school intervention.

## Exhibit 27
## Characteristics of the Participating Schools

| School | Grades | Enrollment | | | | | 2010–2011 Percentages | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Year 1 2006–07 | Year 2 2007–08 | Year 3 2008–09 | Year 4 2009–10 | Year 5 2010–11 | Non-White | F & R Meals | Special Ed | ELL |
| **Middle** | | | | | | | | | | |
| Alpha | 6–8 | 484 | 346 | 236 | 255 | 251 | 72 | 84 | 15 | 33 |
| Beta | 6–8 | 286 | 264 | 197 | 162 | 153 | 86 | 91 | 16 | 42 |
| Gamma[a] | 6–8 | 131 | 159 | 154 | 182 | 206 | 57 | 55 | 19 | — |
| Delta | 6–8 | 471 | 315 | 176 | 178 | 153 | 36 | 40 | 15 | 5 |
| Epsilon[a] | 6–8 | 269 | — | — | — | — | — | — | — | — |
| Zeta | 6–8 | 383 | 328 | 375 | 388 | 364 | 77 | 86 | 22 | 19 |
| Omega | 6–8 | 527 | 489 | 419 | 397 | 398 | 61 | 85 | 20 | 20 |
| **High** | | | | | | | | | | |
| Kappa | 9–12 | 1,283 | 1,233 | 1,007 | 1,032 | 1,036 | 41 | 45 | 14 | 6 |
| Lambda | 9–12 | 566 | 548 | 477 | 435 | 415 | 87 | 75 | 24 | 8 |
| Sigma | 9–12 | 936 | 859 | 900 | 860 | 910 | 61 | 64 | 19 | 14 |
| Theta | 9–12 | 794 | 730 | 703 | 681 | 683 | 70 | 76 | 22 | 9 |

*Note*. F & R Meals = free and reduced-price meals. ELL = English language learner. In 2007–2008, Beta middle school and Delta middle school merged with K–5 schools; Epsilon middle school closed and the final Grade 8 went to Kappa high school for 1 year; Gamma middle school converted from a coed to an all-girls school with new teachers. In 2008–2009, Alpha middle school merged with a K–5 school; Zeta and Omega middle schools were added to the program. Percentages based on whole school enrollment.

[a]Due to lack of postimplementation data, Epsilon middle school was excluded from the impact analysis. The change in composition of Gamma middle school after Year 1 precluded using its data for measuring change and this school was also omitted from the impact analysis.

### *Whole School Intervention Teacher Characteristics*

The evaluation team collected teacher background information through surveys administered to all teachers each year. Exhibits 28 and 29 summarize the characteristics of the whole school intervention teachers. Some notable changes in teacher characteristics occurred from Year 1 to Year 3 as implementation of the whole school intervention was introduced across all subject areas. A greater percentage of teachers reported a Bachelor's as the highest degree obtained in Year 3 in contrast to Year 1 (31% and 20%, respectively), a lower percentage of teachers had a standard or continuing teacher license in Year 3 than in Year 1 (42% and 60%, respectively), and a higher percentage of teachers had an initial teaching license in Year 3 than in Year 1 (35% and 19%, respectively). Teacher characteristics in Years 4 and 5 were similar to Year 3.

**Exhibit 28**
**Experience of the Whole School Intervention Teachers**

| | Year 1 | | Year 2 | | Year 3 | | Year 4 | | Year 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Characteristic** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** |
| **Number of years taught** | 14.4 | 8.8 | 12.8 | 8.7 | 11.9 | 9.2 | 12.6 | 9.1 | 13.2 | 8.2 |
| In Portland Public Schools | 11.5 | 7.9 | 10.4 | 7.6 | 9.5 | 7.9 | 10.1 | 7.9 | 10.9 | 7.8 |
| At current school | 8.2 | 7.1 | 7.7 | 6.4 | 6.6 | 6.3 | 7.3 | 6.3 | 8.0 | 6.6 |

*Note*. Number of teacher respondents = 86 in Year 1, 144 in Year 2, 194 in Year 3, and 156 in Year 4.

**Exhibit 29**
**Characteristics of the Whole School Intervention Teachers**

| Characteristic | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
|---|---|---|---|---|---|
| **Gender** | | | | | |
| Female | 51% | 48% | 58% | 59% | 58% |
| Male | 49% | 52% | 42% | 41% | 42% |
| **School level** | | | | | |
| High school | 60% | 71% | 48% | 51% | 0% |
| Middle school | 40% | 29% | 52% | 49% | 100% |
| **Highest degree obtained** | | | | | |
| BA | 20% | 26% | 31% | 30% | 36% |
| MA | 64% | 63% | 61% | 63% | 53% |
| Education specialist or certification beyond Masters | 9% | 8% | 6% | 5% | 11% |
| Doctorate or professional degree | 4% | 1% | 1% | 1% | 0% |
| Other | 4% | 1% | 1% | 1% | 0% |

| Characteristic | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
|---|---|---|---|---|---|
| **Reading endorsement** | 6% | 4% | 3% | 3% | 4% |
| **Licensure** | | | | | |
| Standard or continuing teaching license | 60% | 54% | 42% | 47% | 43% |
| Basic teaching license | 23% | 19% | 22% | 27% | 32% |
| Initial teaching license | 19% | 27% | 35% | 30% | 24% |
| **Previous experience with content enhancement routines** | 2% | 6% | — | — | — |

*Note*. In Year 1, the number of teacher respondents = 86 (95% of whole school intervention teachers); in Year 2, respondents = 144 (80%); in Year 3, respondents = 194 (83%); in Year 4, respondents = 156 (79%); and in Year 5, respondents = 74 (middle school only). High school teachers were less likely to respond to surveys than middle school teachers. Dashes indicate that question was not asked in those years.

### *Classroom Characteristics*

In Year 1 the whole school intervention involved language arts and social studies teachers only. There were 32 language arts and social studies teachers at the middle school level and 59 at the high school level (total = 91). In Year 2 math and science teachers began implementing the whole school intervention. A total of 84 math and science teachers (26 at the middle school level and 58 at the high school level) and 96 language arts and social studies teachers (34 at the middle school level and 62 at the high school level) were involved in the whole school intervention in Year 2. In Year 3 all teachers in the Striving Reading schools were involved in the whole school intervention. The total population of teachers implementing the whole school intervention in Year 3 was 235 (101 middle school teachers and 134 high school teachers). This total includes 84 language arts and social studies teachers, 83 math and science teachers, 8 self-contained classroom teachers, 22 art and music teachers, 21 physical education and health teachers, and 17 teachers of other subjects (e.g., media, special education, ESL). In Year 4, all teachers in the middle and high schools were expected to participated in the whole school intervention, but in Year 5 only the middle school teachers were expected to participate.

## Impact on Schools

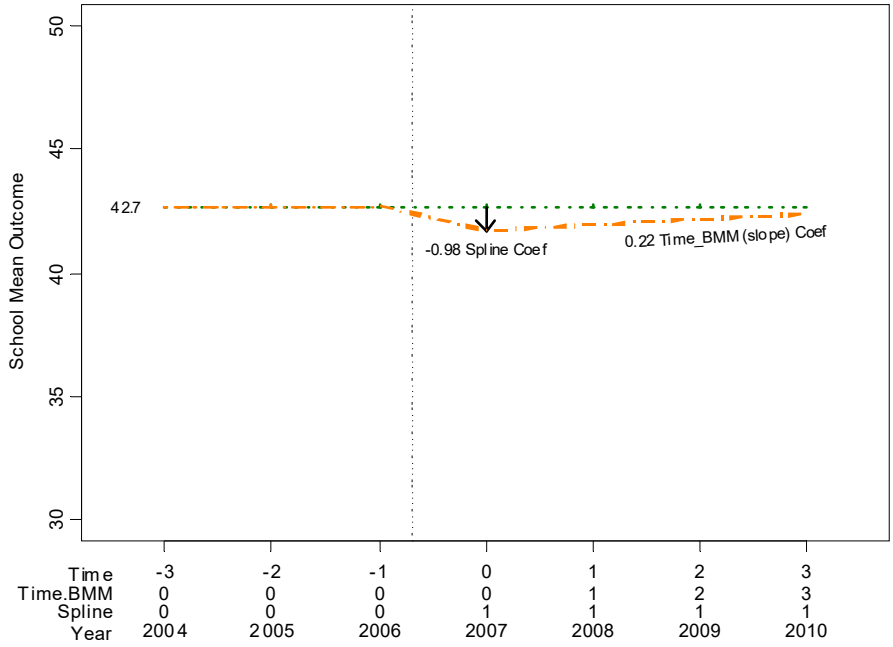
A multilevel model with repeated observations collected on schools before and after implementation of Striving Readers was used to estimate the impact of the whole school intervention. Data included mean school level OAKS total reading NCE scores for 3 years of preimplementation and up to 3 years of postimplementation.[9] An analysis of the pre-

[9]Two schools started 2 years after the initial implementation of Striving Readers, and thus only 2 years of postimplementation data were available for those schools. In Year 5 the changes to the OAKS made scores in Year 5

implementation data revealed that the baseline slope was not significantly different than 0, β = -0.37, *p* = .383. A baseline mean projection model based on data plots and hierarchical linear modeling results was selected for this analysis.

The data analysis revealed no significant effect of the whole school intervention on student total reading NCE scores, β = -1.02, *p* = .235, when controlling for school level, percentage of Hispanic students, and percentage of African American students. The analysis also revealed that the postimplementation slope was significantly different from 0 and in the negative direction, β = 0.22, *p* = .572. Exhibit 30 provides a graphical representation of the intervention effect.

**Exhibit 30**
**Whole School Intervention Effects**




## Discussion

The analysis of the Striving Readers whole school intervention revealed no significant gains in OAKS total reading NCE scores from pre- to postimplementation. This finding was not unexpected for the following reasons: multiple school reconfigurations and staffing changes occurred in the Striving Readers schools during the implementation period; implementation of

---

non-comparable to scores from prior years. Thus Year 5 results are not included in the analysis.

the whole school intervention was weak overall, particularly in the high schools; changes to the OAKS test occurred; and new curricula were introduced during the implementation period. These barriers limited the evaluation team's ability to assess the actual impact of the intervention.

# Appendix A
# Stages of Xtreme Reading Strategy Instruction

## Exhibit A.1
## Stages of Xtreme Reading Strategy Instruction

| Stage | Instructional Practice | Purpose |
|---|---|---|
| 1 | Describing | Teacher provides rationale and describes steps for the strategy. |
| 2 | Teacher Modeling | Teacher demonstrates the strategy by thinking aloud and gradually involving students. |
| 3 | Verbal Practice | Students verbally rehearse the steps of the strategy until they can understand and name the strategy steps. |
| 4 | Guided Practice | Teacher models expert reading behaviors using current and previously learned strategies and prompts students to use strategy steps. |
| 5 | Paired Practice | Students practice the strategy with a peer using materials at their instructional level and provide feedback to each other. Students periodically read to each other, checking accuracy and fluency on timed oral reading passages. The teacher monitors the pairs and provides feedback. |
| 6 | Independent Practice | Students apply the reading strategy to a passage using a worksheet to record their use of the strategy. Students then take a reading comprehension test. |
| 7 | Differentiated Practice | Students apply the reading strategy to individual oral reading with the teacher, and the teacher provides more specific individual feedback (occurs during independent practice time). |
| 8 | Integration and Generalization | Students apply strategies to text from other classes and participate in class discussion of strategy use. |

*Note.* Information provided by the Center for Research on Learning at the University of Kansas, November 2007.

# Appendix B
# Calculation of Implementation Ratings

## Calculation of Implementation Ratings for Targeted Intervention

This section summarizes the rating scales used to quantify implementation of the Targeted Intervention. Three general categories of implementation were rated: professional development, classroom fidelity, and teacher buy-in.

### *Level of Fidelity of Professional Development Model*

#### Year 1

**Group Professional Development (PD) Participation** was based on school district records of attendance at 5 professional development sessions for 2006–2007. Teachers were given a rating from 1 to 4:

4 = participation in 76–100% of the professional development sessions offered
3 = participation in 51–75% of the professional development sessions offered
2 = participation in 26–50% of the professional development sessions offered
1 = participation in 0–25% of the professional development sessions offered

**Teacher Qualification** rating was based on whether the teacher had a reading endorsement:

1 = If the teacher had a reading endorsement
0 = If the teacher did not have a reading endorsement

**Summary Professional Development Scores:**

4–5 = High
3 = Medium
2 = Low
1 = Very Low

#### Year 2

**Group PD Participation** was based on school district records of attendance at 8 professional development sessions for 2007–2008. **Teacher Qualification** ratings and **Summary Professional Development Scores** were given on the same scales listed for Year 1.

#### Year 3

**Group PD Participation** was based on school district records of attendance at professional development sessions offered for 2008–2009. **Teacher Qualification** ratings and **Summary Professional Development Scores** were given on the same scales listed for Year 1.

#### Year 4

**Group PD Participation** was based on school district records of attendance at professional development sessions offered for 2009–2010. **Teacher Qualification** ratings and **Summary Professional Development Scores** were given on the same scales listed for Year 1.

#### Year 5

**Group PD Participation** was based on school district records of attendance at professional development sessions offered for 2010–1011. Veteran Xtreme Reading teachers had no professional development requirements, so Year 4 scores were used. New Xtreme Reading teachers were expected to participate in 8 to 10 sessions. **Teacher Qualification** ratings and **Summary Professional Development Scores** were given on the same scales listed for Year 1.

### *Level of Fidelity of Classroom Model*

#### Year 1

**Classroom Observation** ratings were based on 2 observations of Xtreme lessons per school, generally with 1 observation in winter and 1 in spring. Teachers were rated based on the percentage of expected activities they completed during an Xtreme reading lesson. Percentages for each observation were averaged to determine an implementation fidelity rating for each school:

4 = Fidelity of 90% or higher (High)
3 = Fidelity of 80–89% (Medium)
2 = Fidelity of 70–79% (Low)
1 = Fidelity of less than 70% (Very Low)

**Year 2**

**Classroom Observation** ratings were based on 3 to 5 observations of Xtreme lessons per school during the school year. Teachers were rated based on the percentage of expected activities they completed during an Xtreme reading lesson. Percentages for each observation were averaged to determine an implementation fidelity rating for each school using the same rating scale as Year 1.

**Year 3**

**Classroom Observation** ratings were calculated in the same way as described for Year 2.

**Year 4**

**Classroom Observation** ratings were based on 4 to 8 observations of Xtreme lessons per school during the school year, except for 1 school that permitted only 1 observation. Most teachers were observed 4 times. Percentages for each observation were averaged to determine an implementation fidelity rating for each school using the same rating scale as Year 1.

**Year 5**

**Classroom Observation** ratings were based on 3 to 10 observations of Xtreme lessons per school during the school year. Most teachers were observed 5 times. Percentages for each observation were averaged to determine an implementation fidelity rating for each school using the same rating scale as Year 1.

***Teacher Buy-In***

Teacher buy-in for Years 1 through 5 was calculated using teacher survey ratings of professional development and support and teacher survey ratings of the effectiveness of the Xtreme reading strategies. The number and content of the items varied from year to year but the general approach to calculating teacher buy-in was the same across years. All ratings used a 1 to 5 scale. Summary ratings were assigned to the averages as follows:

4.2–5 = High
3.5–4.1 = Medium
Below 3.5 = Low

## Calculation of Implementation Ratings for Whole School Intervention

This section summarizes the rating scales used to quantify implementation of the Whole School Intervention. Three general categories of implementation were rated: professional development, classroom fidelity, and teacher buy-in.

### *Level of Fidelity of Professional Development Model*

#### Year 1

**Group Professional Development Participation** was based on school district records of attendance: 4 points for summer training, 3 points for fall training, 2 points for October or November trainings, 1 point for attendance at each additional training session on optional content enhancement routines. Subtract 2 points if not trained in all 4 core Routines. Teachers were given a rating from 1 to 4 using the following schedule:

4 = 7 or more points
3 = 6–6.9 points
2 = 4–5.9 points
1 = less than 4 points

**Coaching from Professional Developers** was based on activity logs from the professional developers. Percentiles are based on average number of minutes across all schools. For middle schools, ratings are an average for the whole school; for high schools, ratings on coaching are an average by content area.

4 = Amount of time was at or above 75$^{th}$ percentile for all schools
3 = Amount of time was between 50$^{th}$ and 74$^{th}$ percentile for all schools
2 = Amount of time was between 26$^{th}$ and 49$^{th}$ percentile for all schools
1 = Amount of time was at or below 25$^{th}$ percentile for all schools

**Coaching from School Literacy Coach** was based on activity logs from school literacy coaches. Percentiles are based on average number of minutes across all schools. For middle schools, ratings are an average for the whole school; for high schools, ratings on coaching are an average by content area. Rating scale is the same as for Professional Developer coaching.

**School Literacy Coach Qualifications** were rated based on whether the coach had a Master's Degree and/or a reading endorsement:

4 = Master's degree or higher, *and* Reading endorsement
3 = Master's degree only
2 = Reading endorsement only
1 = Neither a Master's degree or a reading endorsement

**Summary Rating Across 4 Professional Development Areas** took the average rating across the 4 areas:

3.5–4.0 = High
2.5–3.4 = Medium
1.5–2.4 = Low
Below 1.5 = Very Low

**Year 2**

**Group Professional Development Participation** was based on school district records of attendance. 4 points for attendance in June or August 2007 and 3 points for attendance at September 2007 training. 2 points for each training in additional Routines. Subtract 2 points if not trained in all required Routines for content area. Teachers were given a rating from 1 to 4 using the following schedule:

4 = 5 or more points
3 = 4–4.9 points
2 = 3–3.9 points
1 = less than 3 points

**Coaching from Professional Developers** was based on logs from developers. Percentiles are based on average number of minutes across all schools. For middle schools, ratings are an average for the whole school; for high schools, ratings on coaching are an average by content area. Evaluators used the same point schedule as in Year 1.

**Coaching from School Literacy Coach** was based on logs from literacy coaches. Percentiles are based on average number of minutes across all schools. For middle schools, ratings are an

average for the whole school; for high schools, ratings are an average by content area. Evaluators used the same point schedule as in Year 1.

**School Literacy Coach Qualifications** were rated based on whether the coach had a Master's Degree and/or a reading endorsement. Evaluators used the same point schedule as in Year 1.

**Summary Rating Across 4 Professional Development Areas** took the average rating across the 4 areas using the same point schedule as in Year 1.

### Year 3

**Group Professional Development Participation** was based on school district records of attendance: 4 points for attendance in June or August 2008, 3 points for attendance at September 2008 training, and 2 points for each training in additional Routines. Subtract 2 points if not trained in all required Routines for content area. Teachers were given a rating from 1 to 4 using the following schedule:

4 = 5 or more points
3 = 4–4.9 points
2 = 3–3.9 points
1 = less than 3 points

**Coaching from Professional Developers** was based on logs from developers. Percentiles are based on average number of minutes across all schools. For middle schools, ratings are an average for the whole school; for high schools, ratings on coaching are an average by content area. Evaluators used the same point schedule as in Year 1.

**Coaching from School Literacy Coach** was based on logs from literacy coaches. Percentiles are based on average number of minutes across all schools. For middle schools, ratings are an average for the whole school; for high schools, ratings are an average by content area. Evaluators used the same point schedule as in Year 1.

**School Literacy Coach Qualifications** were rated based on whether the coach had a Master's Degree and/or a reading endorsement. Evaluators used the same point schedule as in Year 1.

**Summary Rating Across 4 Professional Development Areas** took the average rating across the 4 areas using the same point schedule as in Year 1.

### Year 4

**Group Professional Development Participation** was based on school district records of attendance. For teachers who had completed basic training, group professional development in Year 4 was optional. The evaluation team gave a bonus point to teachers who participated in June or August 2009 training. Teachers were given a rating from 1 to 4 using the following schedule:

4 = 5 or more points
3 = 4–4.9 points
2 = 3–3.9 points
1 = less than 3 points

**Coaching from Professional Developers** was not tracked for Year 4 because their contract ended midyear at most schools and the focus was on working with school literacy coaches to provide support to content teachers.

**Coaching from School Literacy Coach** was based on logs from literacy coaches. Percentiles are based on average number of minutes across all schools. For middle schools, ratings are an average for the whole school; for high schools, ratings are an average by content area. Evaluators used the same point schedule as in Year 1.

**School Literacy Coach Qualifications** were rated based on whether the coach had a Master's Degree and/or a reading endorsement. Evaluators used the same point schedule as in Year 1.

**Summary Rating Across 3 Professional Development Areas** took the average rating across the 3 areas using the same point schedule as in Year 1. Since the first 3 years of the project used ratings across 4 professional development areas, average scores on professional development and inputs are not exactly comparable between Year 4 and prior years.

### Year 5

**Group Professional Development Participation** was based on school district records of attendance. For teachers who had completed basic training, group professional development in Year 4 was optional. The evaluation team gave a bonus point to teachers who participated in June or August 2009 training. Teachers were given a rating from 1 to 4 using the following schedule:

4 = 5 or more points
3 = 4–4.9 points
2 = 3–3.9 points
1 = less than 3 points

**Coaching from Professional Developers** was not tracked for Year 5 because their contract ended in Year 4.

**Coaching from School Literacy Coach** was not tracked for Year 5 because coaches continued in the 6 middle schools only.

**School Literacy Coach Qualifications** were rated based on whether the coach had a Master's Degree and/or a reading endorsement. Evaluators used the same point schedule as in Year 1.

**Summary Rating Across 2 Professional Development Areas** took the average rating across the 2 areas using the same point schedule as in Year 1. Since the first 3 years of the project used ratings across 4 professional development areas, Year 4 used ratings across 3 professional development areas, and Year 5 used ratings across only 2 professional development areas, average scores on professional development and inputs are not exactly comparable between Year 5 and prior years.

### *Level of Fidelity of Classroom Model*

#### Year 1

**Classroom Observation** ratings were based on classroom observations for teachers who consented to have researchers in their classrooms. Teachers were rated based on the percentage of expected content enhancement routine components they completed during a lesson. Fidelity estimates are more representative of middle school teachers than high school teachers due to small sample sizes from high schools. Percentages for each observation were averaged to determine an implementation rating for each school:

4 = Fidelity of 90% or higher (High)
3 = Fidelity of 80–89% (Medium)
2 = Fidelity of 70–79% (Low)
1 = Fidelity of less than 70% (Very Low)

**Percentage of Required Content Enhancement Routines Completed.** Percentage is based on the average for a school, with data from school literacy coach records. Requirements were determined by the school district, and included a specific number of each type of device. Full-time teachers were expected to complete a total of 15 devices during the school year. Ratings used the following scale:

4 = 75% or higher
3 = 50–74%
2 = 25–49%
1 = Less than 25%

**Percentage of Teachers using Content Enhancement Routines in Previous Month**. Based on teacher survey data, any reported use of a content enhancement routine in the month prior to the survey = 1, no use = 0. An average was created from fall, winter, and spring teacher surveys and rated on the following scale:

4 = 75% or higher
3 = 50–74%
2 = 25–49%
1 = Less than 25%

**Summary Rating Across 3 Classroom Implementation Areas** took the average rating across the 3 areas using the following scale:

3.5–4.0 = High
2.5–3.4 = Medium
1.5–2.4 = Low
Below 1.5 = Very Low

### Year 2

**Classroom Observation** ratings were based on observations during the fall and spring. Typically a higher percentage of middle school teachers than high school teachers participated per school, making the middle school observations more representative. Evaluators used the same point schedule as in Year 1.

**Percentage of Required Content Enhancement Routines Completed** is based on the average for a school, using data from school literacy coach records. Requirements were determined by the school district, and full-time teachers were expected to complete 8 devices during the school year (reduced from Year 1 expectations). In Year 4, requirements were specific to each school, and school literacy coaches determined whether teachers met requirements. Evaluators used the same point schedule as in Year 1.

**Percentage of Teachers Using Content Enhancement Routines in Previous Month**. Based on teacher survey data, any self-reported use of a content enhancement routine in the month prior to the survey = 1, no use = 0. An average was created from fall, winter, and spring teacher surveys and rated on the same point schedule as in Year 1.

**Summary Rating Across 3 Classroom Implementation Areas** took the average rating across the 3 areas using the same point schedule as in Year 1.

#### Year 3

Ratings of classroom implementation were calculated for the 3 areas using the same methods described above for Year 2, except that teacher survey data was collected only in the fall and spring.

#### Year 4

Ratings of classroom implementation were calculated for the 3 areas using the same methods described above for Year 2, except that teacher survey data was collected only in the fall and spring.

#### Year 5

Classroom implementation was not observed in Year 5.

### *Teacher Buy-In*

Teacher buy-in for Years 1 through 5 was calculated using teacher survey ratings of professional development and support and teacher survey ratings of the effectiveness of the content enhancement routines. All ratings used a 1 to 5 scale. Summary ratings were assigned to the averages as follows:

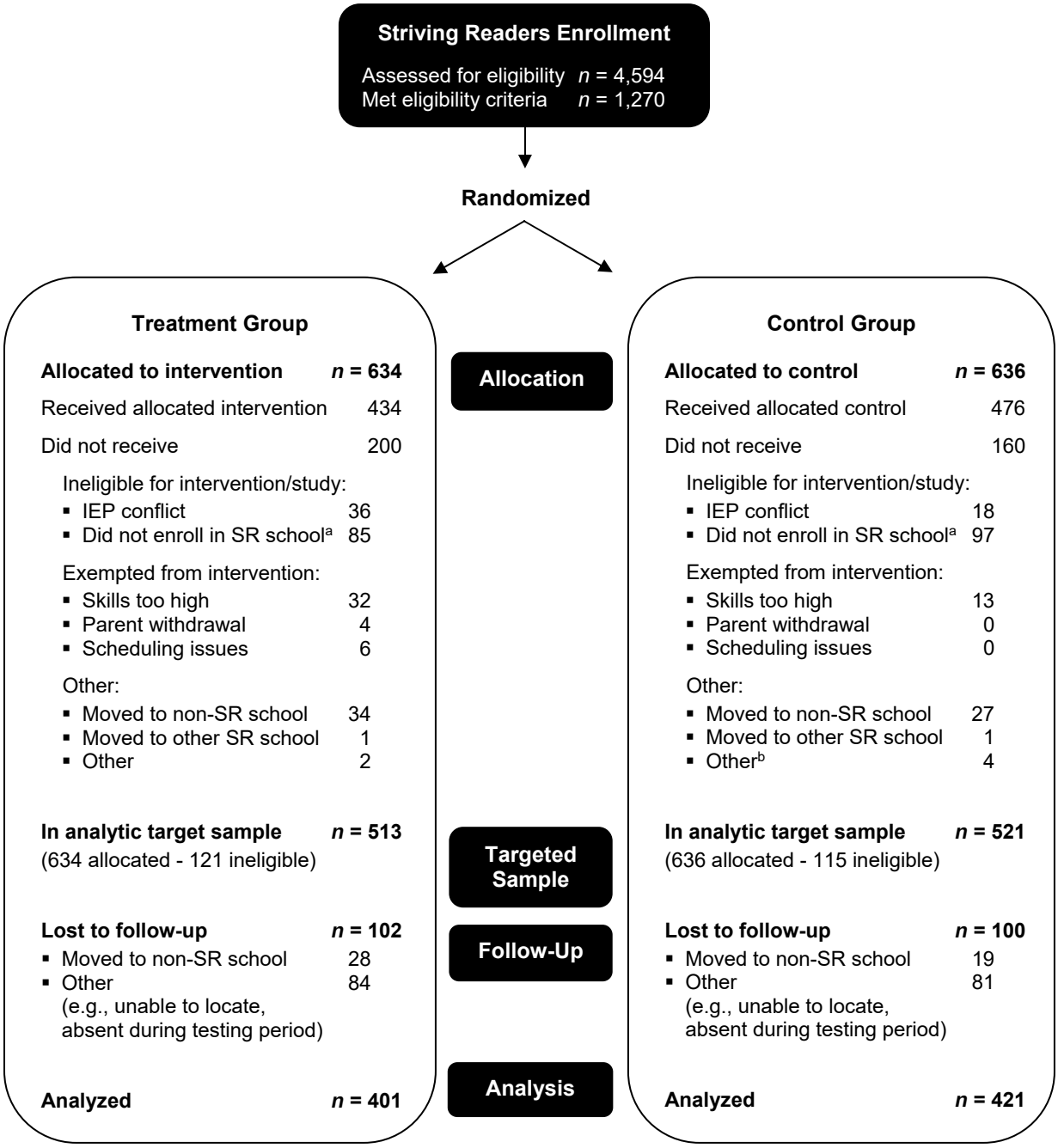
4.2–5 = High

3.5–4.1 = Medium

Below 3.5 = Low

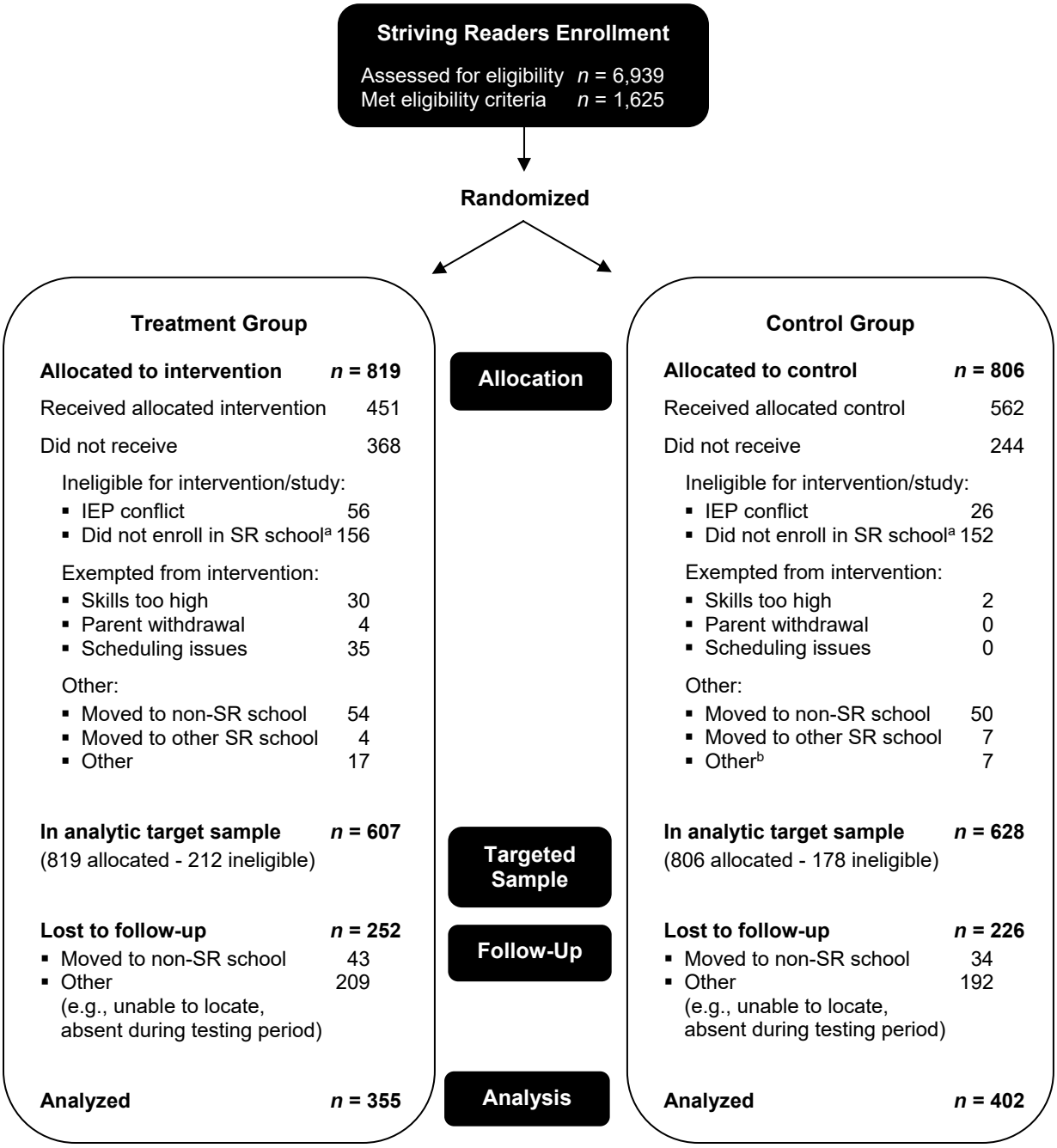
# Appendix C
# GRADE Analytic Samples by School Level

**Exhibit C.1**
**Targeted Intervention Random Assignment for Years 1 through 4**
**Middle School GRADE Sample**




*Note.* IEP = individualized education program, SR = Striving Readers.
[a]These students never learned their group assignment. [b]Crossovers = 4.

**Exhibit C.2**
**Targeted Intervention Random Assignment for Years 1 through 4**
**High School GRADE Sample**




*Note.* IEP = individualized education program, SR = Striving Readers.
[a]These students never learned their group assignment. [b]Crossovers = 5.

# Appendix D
# Baseline Equivalence Tests for GRADE and OAKS Analytic Samples

## Exhibit D.1
## Baseline Equivalence of Treatment and Control Groups on Outcome Measures and Demographic Characteristics

**Baseline Balance Test for Middle School (Years 1 through 4 Combined)**

| | Treatment Group | | | | Control Group | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Baseline Characteristic** | ***n*** | ***n* miss.** | ***M*** | ***SD*** | ***n*** | ***n* miss.** | ***M*** | ***SD*** | ***p*** |
| **GRADE Analysis Sample** | | | | | | | | | |
| GRADE NCE[a, b] | 374 | 27 | 26.94 | 13.54 | 387 | 34 | 24.95 | 15.11 | .055 |
| OAKS NCE[a, b] | 390 | 11 | 27.42 | 11.41 | 399 | 22 | 26.81 | 12.08 | .464 |
| Motivation to Read (Cohort 1)[a, c] | 128 | 36 | 2.75 | 0.54 | 147 | 33 | 2.72 | 0.55 | .585 |
| Motivation to Read (Cohort 2)[a, d] | 25 | 20 | 3.75 | 0.91 | 40 | 14 | 3.42 | 0.92 | .164 |
| Motivation to Read (Cohort 3)[a, d] | 83 | 27 | 3.30 | 0.82 | 90 | 17 | 3.59 | 1.01 | .040 |
| Motivation to Read (Cohort 4)[a, d] | 78 | 4 | 3.57 | 0.95 | 79 | 1 | 3.55 | 0.97 | .886 |
| Gender | 401 | 0 | 0.47 | 0.50 | 421 | 0 | 0.47 | 0.50 | .866 |
| Black | 401 | 0 | 0.23 | 0.42 | 421 | 0 | 0.23 | 0.42 | .895 |
| Hispanic | 401 | 0 | 0.31 | 0.46 | 421 | 0 | 0.36 | 0.48 | .136 |
| SPED Status | 401 | 0 | 0.28 | 0.45 | 421 | 0 | 0.26 | 0.44 | .512 |
| ELL Status | 401 | 0 | 0.36 | 0.48 | 421 | 0 | 0.32 | 0.47 | .246 |
| **OAKS Analysis Sample** | | | | | | | | | |
| GRADE NCE[a, b] | 421 | 51 | 27.03 | 13.54 | 427 | 55 | 24.84 | 14.97 | .026 |
| OAKS NCE[a, b] | 455 | 17 | 27.35 | 11.50 | 455 | 27 | 26.78 | 12.10 | .470 |
| Motivation to Read (Cohort 1)[a, c] | 130 | 68 | 2.75 | 0.54 | 152 | 61 | 2.71 | 0.55 | .522 |
| Motivation to Read (Cohort 2)[a, d] | 25 | 34 | 3.75 | 0.91 | 42 | 18 | 3.38 | 0.94 | .128 |
| Motivation to Read (Cohort 3)[a, d] | 91 | 33 | 3.34 | 0.82 | 98 | 22 | 3.56 | 1.02 | .107 |
| Motivation to Read (Cohort 4)[a, d] | 84 | 7 | 3.60 | 0.95 | 86 | 3 | 3.50 | 0.98 | .503 |
| Gender | 472 | 0 | 0.48 | 0.50 | 482 | 0 | 0.48 | 0.50 | .988 |
| Black | 472 | 0 | 0.26 | 0.44 | 482 | 0 | 0.23 | 0.42 | .246 |
| Hispanic | 472 | 0 | 0.31 | 0.46 | 482 | 0 | 0.36 | 0.48 | .092 |
| SPED Status | 472 | 0 | 0.26 | 0.44 | 482 | 0 | 0.26 | 0.44 | .907 |
| ELL Status | 472 | 0 | 0.34 | 0.47 | 482 | 0 | 0.31 | 0.46 | .398 |

[a]Pretest. [b]NCE scores range from 1 to 99. [c]Cohort 1 Motivation to Read scores range from 1 to 4 (1 = *very different from me, 2 = a little different from me, 3 = a little like me,* 4 = *a lot like me).* [d]Cohorts 2 and 3 Motivation to Read scores range from 1 to 6 (1 = *strongly agree,* 6 = *strongly disagree).*

**Baseline Balance Test for High School (Years 1 through 4 Combined)**

| | Treatment Group | | | | Control Group | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Baseline Characteristic** | ***n*** | ***n* miss.** | ***M*** | ***SD*** | ***n*** | ***n* miss.** | ***M*** | ***SD*** | ***p*** |
| **GRADE Analysis Sample** | | | | | | | | | |
| GRADE NCE[a, b] | 303 | 52 | 30.20 | 16.10 | 336 | 66 | 28.73 | 15.76 | .243 |
| OAKS NCE[a, b] | 316 | 39 | 28.71 | 12.32 | 366 | 36 | 28.12 | 12.95 | .544 |
| Motivation to Read (Cohort 1)[a, c] | 76 | 56 | 2.69 | 0.50 | 81 | 73 | 2.74 | 0.55 | .517 |
| Motivation to Read (Cohort 2)[a, d] | 26 | 16 | 3.85 | 0.92 | 63 | 15 | 3.43 | 0.92 | .054 |
| Motivation to Read (Cohort 3)[a, d] | 53 | 53 | 3.33 | 0.89 | 79 | 22 | 3.24 | 1.06 | .595 |
| Motivation to Read (Cohort 4)[a, d] | 40 | 35 | 2.94 | 1.11 | 64 | 5 | 3.50 | 0.87 | .005 |
| Gender | 355 | 0 | 0.48 | 0.50 | 402 | 0 | 0.48 | 0.50 | .918 |
| Black | 355 | 0 | 0.34 | 0.48 | 402 | 0 | 0.36 | 0.48 | .576 |
| Hispanic | 355 | 0 | 0.26 | 0.44 | 402 | 0 | 0.25 | 0.43 | .736 |
| SPED Status | 355 | 0 | 0.21 | 0.41 | 402 | 0 | 0.27 | 0.45 | .054 |
| ELL Status | 355 | 0 | 0.20 | 0.40 | 402 | 0 | 0.19 | 0.39 | .712 |
| **OAKS Analysis Sample** | | | | | | | | | |
| GRADE NCE[a, b] | 162 | 98 | 32.19 | 16.90 | 182 | 72 | 31.49 | 16.24 | .695 |
| OAKS NCE[a, b] | 200 | 60 | 32.89 | 12.70 | 198 | 56 | 33.21 | 14.93 | .818 |
| Motivation to Read (Cohort 1)[a, c] | 39 | 85 | 2.66 | 0.54 | 38 | 79 | 2.75 | 0.50 | .457 |
| Motivation to Read (Cohort 2)[a, d] | 13 | 33 | 3.78 | 1.05 | 27 | 16 | 3.58 | 0.84 | .517 |
| Motivation to Read (Cohort 3)[a, d] | 22 | 38 | 3.31 | 0.89 | 44 | 17 | 3.34 | 1.05 | .896 |
| Motivation to Read (Cohort 4)[a, d] | 12 | 18 | 3.25 | 1.05 | 28 | 5 | 3.53 | 0.88 | .381 |
| Gender | 260 | 0 | 0.52 | 0.50 | 254 | 0 | 0.51 | 0.50 | .936 |
| Black | 260 | 0 | 0.40 | 0.49 | 254 | 0 | 0.39 | 0.49 | .677 |
| Hispanic | 260 | 0 | 0.20 | 0.40 | 254 | 0 | 0.24 | 0.43 | .271 |
| SPED Status | 260 | 0 | 0.15 | 0.35 | 254 | 0 | 0.20 | 0.40 | .128 |
| ELL Status | 260 | 0 | 0.15 | 0.35 | 254 | 0 | 0.15 | 0.35 | .988 |

[a]Pretest. [b]NCE scores range from 1 to 99. [c]Cohort 1 Motivation to Read scores range from 1 to 4 (1 = *very different from me, 2 = a little different from me, 3 = a little like me,* 4 = *a lot like me).* [d]Cohorts 2 and 3 Motivation to Read scores range from 1 to 6 (1 = *strongly agree,* 6 = *strongly disagree).*

## Exhibit D.2
## Baseline Equivalence of Analytic and Lost to Follow-Up Samples

**Middle School (Years 1 through 4 Combined)**

| | Analysis Sample (Treatment & Control) | | | | Lost to Follow-Up (Treatment & Control) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Baseline Characteristic** | ***n*** | ***n* miss.** | ***M*** | ***SD*** | ***n*** | ***n* miss.** | ***M*** | ***SD*** | ***p*** |
| **GRADE Analysis Sample** | | | | | | | | | |
| GRADE NCE[a, b] | 761 | 61 | 25.93 | 14.39 | 129 | 83 | 25.60 | 14.16 | .809 |
| OAKS NCE[a, b] | 789 | 33 | 27.11 | 11.75 | 195 | 17 | 26.52 | 11.58 | .527 |
| Motivation to Read (Cohort 1)[a, c] | 275 | 69 | 2.73 | 0.54 | 8 | 87 | 2.63 | 0.55 | .601 |
| Motivation to Read (Cohort 2)[a, d] | 65 | 34 | 3.54 | 0.92 | 4 | 27 | 2.95 | 1.10 | .219 |
| Motivation to Read (Cohort 3)[a, d] | 173 | 44 | 3.45 | 0.93 | 28 | 33 | 3.39 | 0.86 | .733 |
| Motivation to Read (Cohort 4)[a, d] | 157 | 5 | 3.56 | 0.96 | 17 | 8 | 3.35 | 1.01 | .390 |
| Gender | 822 | 0 | 0.47 | 0.50 | 211 | 1 | 0.53 | 0.50 | .152 |
| Black | 822 | 0 | 0.23 | 0.42 | 212 | 0 | 0.34 | 0.48 | .002 |
| Hispanic | 822 | 0 | 0.34 | 0.47 | 212 | 0 | 0.28 | 0.45 | .109 |
| SPED Status | 822 | 0 | 0.27 | 0.44 | 212 | 0 | 0.21 | 0.41 | .092 |
| ELL Status | 822 | 0 | 0.34 | 0.47 | 212 | 0 | 0.19 | 0.40 | .000 |
| **OAKS Analysis Sample** | | | | | | | | | |
| GRADE NCE[a, b] | 848 | 106 | 25.93 | 14.31 | 42 | 38 | 24.93 | 15.15 | .669 |
| OAKS NCE[a, b] | 910 | 44 | 27.07 | 11.80 | 74 | 6 | 26.13 | 10.59 | .509 |
| Motivation to Read (Cohort 1)[a, c] | 282 | 129 | 2.73 | 0.54 | 1 | 27 | 3.26 | — | — |
| Motivation to Read (Cohort 2)[a, d] | 67 | 52 | 3.52 | 0.94 | 2 | 9 | 3.15 | 1.09 | .590 |
| Motivation to Read (Cohort 3)[a, d] | 189 | 55 | 3.46 | 0.93 | 12 | 22 | 3.25 | 0.63 | .456 |
| Motivation to Read (Cohort 4)[a, d] | 170 | 10 | 3.55 | 0.97 | 4 | 3 | 3.20 | 0.72 | .478 |
| Gender | 954 | 0 | 0.48 | 0.50 | 79 | 1 | 0.54 | 0.50 | .250 |
| Black | 954 | 0 | 0.25 | 0.43 | 80 | 0 | 0.31 | 0.47 | .238 |
| Hispanic | 954 | 0 | 0.34 | 0.47 | 80 | 0 | 0.21 | 0.41 | .011 |
| SPED Status | 954 | 0 | 0.26 | 0.44 | 80 | 0 | 0.21 | 0.41 | .338 |
| ELL Status | 954 | 0 | 0.32 | 0.47 | 80 | 0 | 0.11 | 0.32 | .000 |

[a]Pretest. [b]NCE scores range from 1 to 99. [c]Cohort 1 Motivation to Read scores range from 1 to 4 (1 = *very different from me, 2 = a little different from me, 3 = a little like me,* 4 = *a lot like me).* [d]Cohorts 2 and 3 Motivation to Read scores range from 1 to 6 (1 = *strongly agree,* 6 = *strongly disagree).*

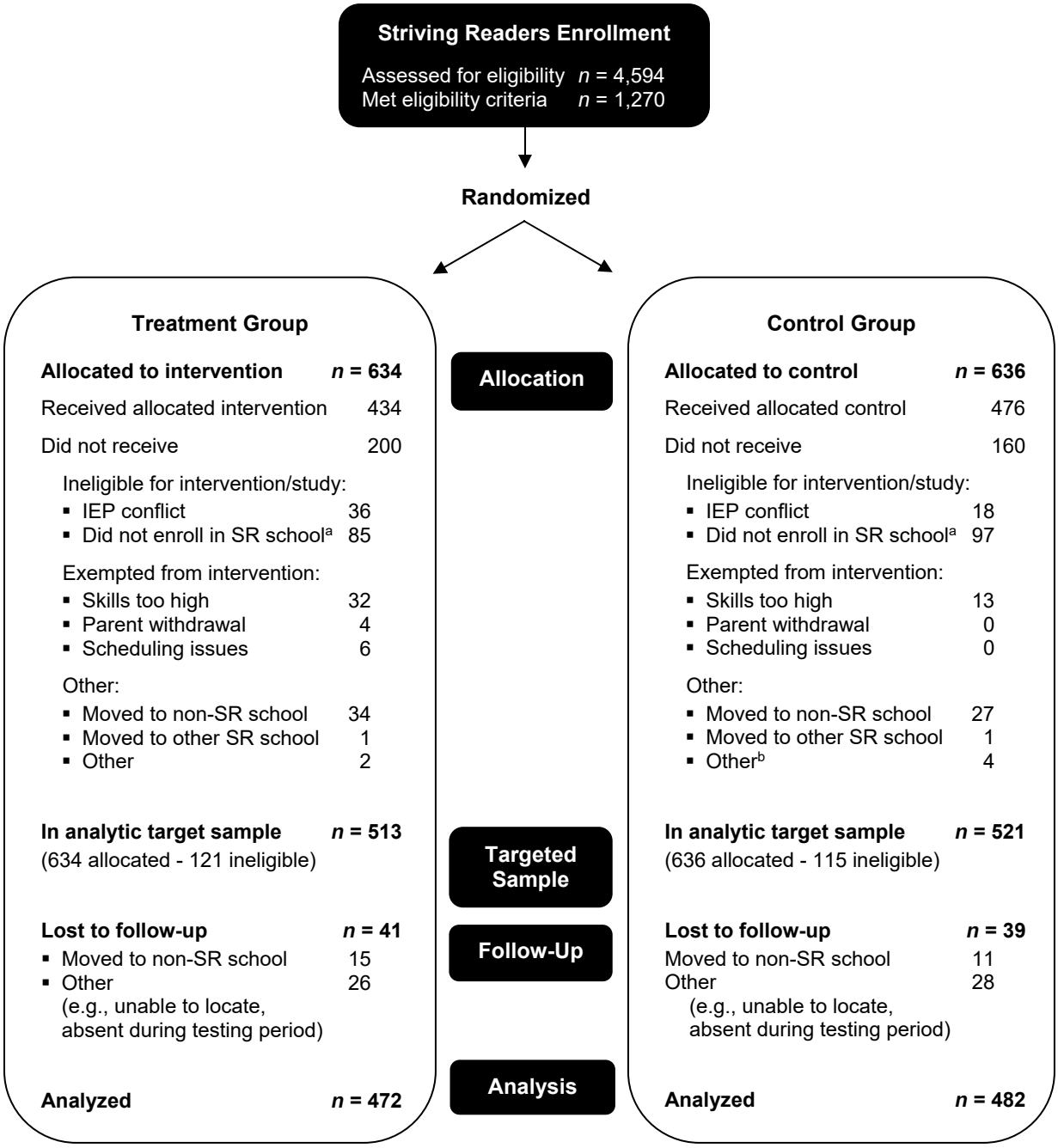
**High School (Years 1 through 4 Combined)**

| Baseline Characteristic | Analysis Sample (Treatment & Control) | | | | Lost to Follow-Up (Treatment & Control) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *n* | *n* miss. | *M* | *SD* | *n* | *n* miss. | *M* | *SD* | *p* |
| **GRADE Analysis Sample** | | | | | | | | | |
| GRADE NCE[a, b] | 639 | 118 | 29.43 | 15.93 | 242 | 236 | 29.42 | 16.76 | .993 |
| OAKS NCE[a, b] | 682 | 75 | 28.40 | 12.66 | 383 | 95 | 27.09 | 13.45 | .122 |
| Motivation to Read (Cohort 1)[a, c] | 157 | 129 | 2.71 | 0.53 | 18 | 223 | 2.45 | 0.65 | .051 |
| Motivation to Read (Cohort 2)[a, d] | 89 | 31 | 3.55 | 0.93 | 2 | 83 | 3.81 | 1.41 | .701 |
| Motivation to Read (Cohort 3)[a, d] | 132 | 75 | 3.27 | 0.99 | 29 | 60 | 3.08 | 0.86 | .326 |
| Motivation to Read (Cohort 4)[a, d] | 104 | 40 | 3.28 | 1.00 | 34 | 29 | 3.52 | 1.05 | .244 |
| Gender | 757 | 0 | 0.48 | 0.50 | 476 | 2 | 0.48 | 0.50 | .943 |
| Black | 757 | 0 | 0.35 | 0.48 | 477 | 1 | 0.37 | 0.48 | .496 |
| Hispanic | 757 | 0 | 0.26 | 0.44 | 477 | 1 | 0.24 | 0.43 | .606 |
| SPED Status | 757 | 0 | 0.24 | 0.43 | 477 | 1 | 0.18 | 0.38 | .008 |
| ELL Status | 757 | 0 | 0.19 | 0.39 | 477 | 1 | 0.17 | 0.38 | .337 |
| **OAKS Analysis Sample** | | | | | | | | | |
| GRADE NCE[a, b] | 344 | 170 | 31.82 | 16.53 | 24 | 39 | 29.33 | 20.29 | .484 |
| OAKS NCE[a, b] | 398 | 116 | 33.05 | 13.84 | 29 | 34 | 28.69 | 14.53 | .103 |
| Motivation to Read (Cohort 1)[a, c] | 77 | 164 | 2.71 | 0.52 | 0 | 28 | — | — | — |
| Motivation to Read (Cohort 2)[a, d] | 40 | 49 | 3.64 | 0.91 | 1 | 10 | 2.96 | — | — |
| Motivation to Read (Cohort 3)[a, d] | 66 | 55 | 3.33 | 0.99 | 5 | 29 | 3.33 | 0.81 | .994 |
| Motivation to Read (Cohort 4)[a, d] | 40 | 23 | 3.44 | 0.93 | 5 | 2 | 3.34 | 1.51 | .891 |
| Gender | 514 | 0 | 0.51 | 0.50 | 61 | 2 | 0.44 | 0.50 | .299 |
| Black | 514 | 0 | 0.39 | 0.49 | 62 | 1 | 0.34 | 0.48 | .385 |
| Hispanic | 514 | 0 | 0.22 | 0.41 | 62 | 1 | 0.19 | 0.40 | .685 |
| SPED Status | 514 | 0 | 0.17 | 0.38 | 62 | 1 | 0.15 | 0.36 | .605 |
| ELL Status | 514 | 0 | 0.15 | 0.35 | 62 | 1 | 0.10 | 0.30 | .233 |

[a]Pretest. [b]NCE scores range from 1 to 99. [c]Cohort 1 Motivation to Read scores range from 1 to 4 (1 = *very different from me, 2 = a little different from me, 3 = a little like me,* 4 = *a lot like me).* [d]Cohorts 2 and 3 Motivation to Read scores range from 1 to 6 (1 = *strongly agree, 6 = strongly disagree).*

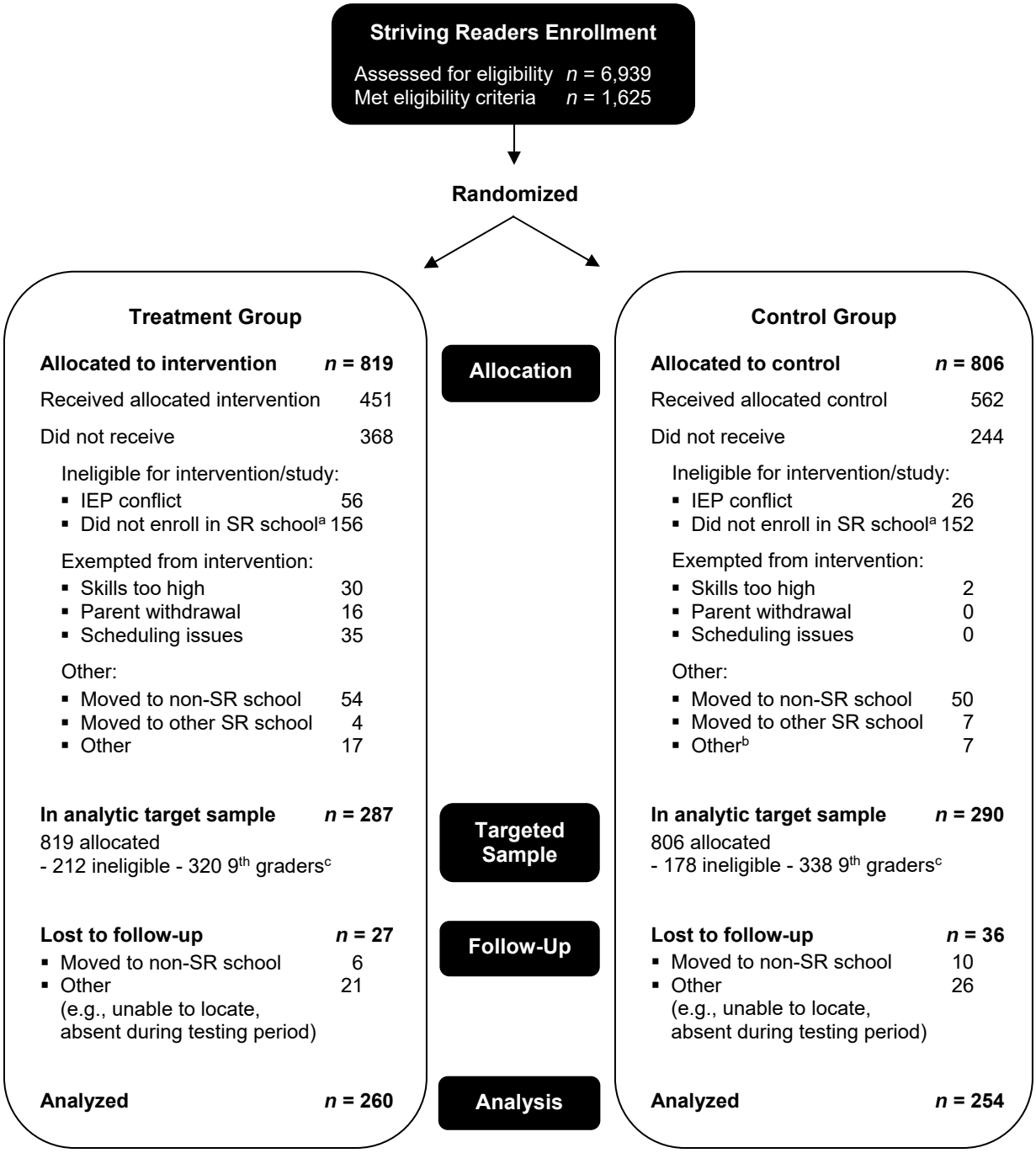
# Appendix E
# OAKS Analytic Samples by School Level

**Exhibit E.1**
**Targeted Intervention Random Assignment for Years 1 through 4**
**Middle School OAKS Sample**




*Note.* IEP = individualized education program, SR = Striving Readers.
[a]These students never learned their group assignment. [b]Crossovers = 4.

**Exhibit E.2**
**Targeted Intervention Random Assignment for Years 1 through 4**
**High School OAKS Sample**




*Note.* IEP = individualized education program, SR = Striving Readers.
[a]These students never learned their group assignment. [b]Crossovers = 5. [c]Students in Grade 9 do not take the OAKS and were removed from the analytic sample.

# Appendix F
# Cross-tabulations of GRADE and OAKS Samples

## Exhibit F.1
## Cross-Tabulations Between GRADE and OAKS
## Analytic Samples

**Middle School Years 1 through 4 Combined**

| | | OAKS at Posttest | | |
|---|---|---|---|---|
| **Group** | | **N** | **Y** | **Totals** |
| **Treatment** | | | | |
| Grade at posttest | N | 41 | 71 | 41 + 71 = 112 |
| | Y | 0 | 401 | 0 + 401 = 401 |
| Total | | 41 | 472 | 41 + 472 = 513 |
| **Control** | | | | |
| Grade at posttest | N | 37 | 63 | 37 + 63 = 100 |
| | Y | 2 | 419 | 2 + 419 = 421 |
| Total | | 39 | 482 | 39 + 482 = 521 |

**High School Years 1 through 4 Combined**

| | | OAKS at Posttest | | |
|---|---|---|---|---|
| **Group** | | **N[a]** | **Y** | **Totals** |
| **Treatment** | | | | |
| Grade at posttest | N | 140 | 112 | 140 + 112 = 252 |
| | Y | 207 | 149 | 207 + 148 = 355 |
| Total | | 347 | 260 | 347 + 260 = 607 |
| **Control** | | | | |
| Grade at posttest | N | 149 | 77 | 149 + 77 = 226 |
| | Y | 225 | 177 | 225 + 177 = 402 |
| Total | | 374 | 254 | 374 + 254 = 628 |

*Note:* Students in Grade 9 do not take the OAKS.

# Appendix G
# GRADE Subtest Outcomes

The same multilevel model used to assess the intervention impact on GRADE Total NCE scores was applied to assess impacts on GRADE Vocabulary and Comprehension NCE scores.

## GRADE Vocabulary Subtest

Results for Vocabulary NCE were consistent with results for Total NCE. For the overall sample, the treatment group students had significantly higher outcomes than the control group students, $\beta = 2.18$, $p < .001$. Whereas a significant treatment effect was present for the middle school sample, $\beta = 3.14$, $p < .001$, there was no significant treatment effect for the high school sample, $\beta = 1.12$, $p = .254$. Exhibit G.1 presents the Vocabulary NCE impact estimates.

**Exhibit G.1**
**GRADE Vocabulary NCE Impact Estimates**

| | | Estimated Impact | | | |
|---|---|---|---|---|---|
| **Group** | ***n*** | **Impact (β)** | ***SE*** | **Effect Size** | ***p*** |
| Middle school | 822 | 3.14 | 0.83 | .20 | .000 |
| High school | 757 | 1.12 | 0.98 | .07 | .254 |
| Overall | 1,579 | 2.18 | 0.64 | .14 | .001 |

Using Glass's Δ method (with the impact estimate as the numerator and the control group standard deviation as the denominator), the standardized effect sizes for the overall sample, middle school sample, and high school sample were .14, .20, and .07, respectively. Exhibit G.2 displays GRADE descriptive statistics in NCEs on the Vocabulary subtest.

**Exhibit G.2**
**GRADE Vocabulary Descriptive Statistics in NCEs**

| | | Middle School | | | High School | | | Overall | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Testing** | **Group** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** |
| Pretest | Treatment | 26.83 | 15.82 | 1–82 | 31.89 | 16.44 | 1–84 | 29.10 | 16.29 | 1–84 |
| | Control | 25.79 | 16.91 | 1–78 | 31.15 | 16.09 | 1–88 | 28.28 | 16.73 | 1–88 |
| Posttest | Treatment | 31.09 | 15.67 | 1–81 | 32.97 | 17.88 | 1–99 | 31.97 | 16.76 | 1–99 |
| | Control | 27.22 | 15.78 | 1–71 | 31.62 | 16.24 | 1–81 | 29.37 | 16.14 | 1–81 |

*Note.* Posttest treatment total $n = 756$; Middle school total $n = 401$; High school total $n = 355$. Posttest control total $n = 823$; Middle school total $n = 421$; High school total $n = 402$.

## GRADE Comprehension Subtest

Results for the Comprehension subtest were consistent with Vocabulary subtest results, although effects were slightly stronger. For the overall sample, the treatment group students had significantly higher outcomes than the control group students, $\beta = 3.96$, $p < .001$. Similarly, a significant treatment effect was present for the middle school sample, $\beta = 5.03$, $p < .001$, and high school sample, $\beta = 2.64$, $p < .05$. Exhibit G.3 presents the impact estimates for the overall sample and by school level.

**Exhibit G.3**
**GRADE Comprehension NCE Impact Estimates**

| | | Estimated Impact | | | |
|---|---|---|---|---|---|
| **Group** | ***n*** | **Impact (β)** | ***SE*** | **Effect Size** | ***p*** |
| Middle school | 822 | 5.03 | 0.79 | .32 | .000 |
| High school | 757 | 2.64 | 1.12 | .15 | .019 |
| Overall | 1,579 | 3.96 | 0.68 | .23 | .000 |

Using Glass's Δ method (with the impact estimate as the numerator and the control group standard deviation as the denominator) to calculate the standardized effect size, the standardized effect sizes for the overall sample, middle school sample, and high school sample were .23, .32, and .15, respectively. Exhibit G.4 displays GRADE descriptive statistics (means, standard deviations, and range) in NCEs on the Comprehension subtest for the overall sample and by school level.

**Exhibit G.4**
**GRADE Comprehension Descriptive Statistics in NCEs**

| | | Middle School | | | High School | | | Overall | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Testing** | **Group** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** |
| Pretest | Treatment | 28.06 | 13.53 | 1–67 | 29.06 | 16.20 | 1–71 | 28.51 | 14.78 | 1–71 |
| | Control | 25.91 | 14.62 | 1–68 | 26.95 | 15.99 | 1–75 | 26.39 | 15.27 | 1–75 |
| Posttest | Treatment | 35.43 | 14.39 | 1–74 | 32.12 | 20.04 | 1–84 | 33.88 | 17.34 | 1–84 |
| | Control | 28.86 | 15.69 | 1–72 | 28.22 | 18.18 | 1–75 | 28.54 | 16.94 | 1–75 |

*Note.* Posttest treatment total $n = 756$; Middle school total $n = 401$; High school total $n = 355$. Posttest control total $n = 823$; Middle school total $n = 421$; High school total $n = 402$.

# Appendix H
# OAKS Outcomes Using GRADE Analytic Samples

## OAKS NCE Outcomes

The same multilevel model used in the main GRADE impact analysis was used to estimate the impact of the intervention on spring OAKS NCEs. The data analysis revealed an intervention effect for the overall sample, $\beta = 1.50$, $p < .05$, and for the middle school sample, $\beta = 2.00$, $p < .01$; however, no significant treatment effect was present for the high school sample. Exhibit H.1 presents the impact estimates for the overall sample and by school level. Glass's Δ method (with the impact estimate as the numerator) was used to calculate the standardized effect sizes.

**Exhibit H.1**
**OAKS NCE Impact Estimates (GRADE Analytic Sample)**

| | Estimated Impact | | | | |
|---|---|---|---|---|---|
| **Group** | ***n*** | **Impact (β)** | ***SE*** | **Effect Size** | ***p*** |
| Middle school | 820 | 2.00 | 0.76 | .15 | .009 |
| High school | 325 | 0.30 | 1.31 | .02 | .817 |
| Overall | 1,145 | 1.50 | 0.66 | .11 | .024 |

Exhibit H.2 displays OAKS descriptives in NCEs for the overall sample and by school level.

**Exhibit H.2**
**OAKS Descriptive Statistics in NCEs (GRADE Analytic Sample)**

| | | Middle School | | | High School | | | Overall | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Testing** | **Group** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** |
| Pretest | Treatment | 27.42 | 11.41 | 1–63 | 32.65 | 13.14 | 1–73 | 28.62 | 12.02 | 1–73 |
| | Control | 26.85 | 12.08 | 1–81 | 32.55 | 15.21 | 1–66 | 28.37 | 13.21 | 1–81 |
| Posttest | Treatment | 32.95 | 12.63 | 1–77 | 35.93 | 12.64 | 1–66 | 33.76 | 12.69 | 1–77 |
| | Control | 30.90 | 13.19 | 1–72 | 35.93 | 13.46 | 1–71 | 32.40 | 13.45 | 1–72 |

*Note.* Posttest treatment total $n = 549$; Middle school total $n = 401$; High school total $n = 148$. Posttest control total $n = 596$; Middle school total $n = 419$; High school total $n = 177$.

To assess whether students missing pretest scores differed from students not missing pretest scores in terms of the OAKS impact, the model controlled for the presence of pretest OAKS NCE scores (present/missing). There were no differences in outcome scores between students who were missing pretest scores and students who were not missing pretest scores.

# Appendix I
# GRADE Outcomes Using GRADE Analytic ELL Sample

## GRADE NCE Outcomes in ELL Sample

The same multilevel model used in the main GRADE impact analysis was used to estimate the impact of the intervention on spring GRADE NCEs in an analytic subsample comprising ELL students. The data analysis revealed an intervention effect for the ELL subsample, $\beta = 4.89$, $p < .001$. Exhibit I.1 presents the impact estimates for the ELL subsample. Glass's Δ method (with the impact estimate as the numerator) was used to calculate the standardized effect size.

**Exhibit I.1**
**OAKS NCE Impact Estimates (GRADE Analytic Sample)**

| | Estimated Impact | | | | |
|---|---|---|---|---|---|
| **Group** | ***n*** | **Impact (β)** | ***SE*** | **Effect Size** | ***p*** |
| ELL subsample | 422 | 4.89 | 1.08 | 0.35 | .000 |

Exhibit I.2 displays GRADE descriptives in NCEs for the GRADE analytic ELL subsample.

**Exhibit I.2**
**GRADE Descriptive Statistics in NCEs (GRADE Analytic ELL Subsample)**

| Testing | Group | *M* | *SD* | Range |
|---|---|---|---|---|
| Pretest | Treatment | 22.41 | 12.39 | 1–49 |
| | Control | 18.75 | 14.00 | 1–64 |
| Posttest | Treatment | 27.84 | 14.21 | 1–57 |
| | Control | 20.18 | 14.13 | 1–64 |

*Note.* Posttest treatment total *n* = 213. Posttest control total *n* = 209.

To assess whether students missing pretest scores differed from students not missing pretest scores in terms of the GRADE impact, the model controlled for the presence of pretest GRADE NCE scores (present/missing). There were no differences in outcome scores between ELL students who were missing pretest scores and students who were not missing pretest scores.

# Appendix J
# GRADE Outcomes Using GRADE Analytic SPED Sample

## GRADE NCE Outcomes in SPED Sample

The same multilevel model used in the main GRADE impact analysis was used to estimate the impact of the intervention on spring GRADE NCEs in an analytic subsample comprising SPED students. The data analysis revealed an intervention effect for the SPED subsample, $\beta = 3.40$, $p < .01$. Exhibit J.1 presents the impact estimates for the SPED subsample. Glass's Δ method (with the impact estimate as the numerator) was used to calculate the standardized effect size.

**Exhibit J.1**
**OAKS NCE Impact Estimates (GRADE Analytic Sample)**

| | Estimated Impact | | | | |
|---|---|---|---|---|---|
| **Group** | ***n*** | **Impact (β)** | ***SE*** | **Effect Size** | ***p*** |
| SPED subsample | 403 | 3.40 | 1.21 | 0.22 | .006 |

Exhibit J.2 displays GRADE descriptives in NCEs for the GRADE analytic SPED subsample.

**Exhibit J.2**
**GRADE Descriptive Statistics in NCEs (GRADE Analytic SPED Subsample)**

| Testing | Group | *M* | *SD* | Range |
|---|---|---|---|---|
| Pretest | Treatment | 21.58 | 13.31 | 1–58 |
| | Control | 21.32 | 15.27 | 1–68 |
| Posttest | Treatment | 27.17 | 17.21 | 1–70 |
| | Control | 23.42 | 15.65 | 1–67 |

*Note.* Posttest treatment total *n* = 186. Posttest control total *n* = 217.

To assess whether students missing pretest scores differed from students not missing pretest scores in terms of the GRADE impact, the model controlled for the presence of pretest GRADE NCE scores (present/missing). There were no differences in outcome scores between SPED students who were missing pretest scores and students who were not missing pretest scores.

# Appendix K
# OAKS Outcomes: Second Year Follow-up

## OAKS NCE Outcomes: Second Year Follow-Up

Second year follow-up data were analyzed to assess whether effects of the intervention were sustained after 1 year postintervention. The same multilevel model used in the main OAKS impact analysis was used to estimate the impact of the intervention on Year 2 follow-up spring OAKS NCE scores.

All Grade 7 students in the OAKS sample with a second year follow-up OAKS score were included in the analysis (*n* = 313). Because the OAKS is not administered to students in Grade 9 or 11, this analysis was not possible for students in Grade 8 (due to missing Grade 9 second year follow-up scores), Grade 9 (due to missing Grade 9 intervention year follow-up scores), or Grade 10 (due to missing Grade 11 second year follow-up scores). Exhibit K.1 shows the OAKS mean, standard deviation, and range by group for each pre and follow-up test.

**Exhibit K.1**
**OAKS Descriptive Statistics in NCEs (OAKS Analytic Sample)**

| Testing | Group | *M* | *SD* | Range |
|---|---|---|---|---|
| Pretest | Treatment | 28.28 | 11.13 | 1–67 |
| | Control | 25.14 | 12.51 | 1–55 |
| Posttest (Year 1) | Treatment | 34.37 | 12.31 | 10–77 |
| | Control | 31.62 | 11.34 | 1–55 |
| Posttest (Year 2) | Treatment | 32.37 | 14.30 | 1–73 |
| | Control | 31.64 | 12.92 | 1–62 |

*Note.* Second year posttest treatment total *n* = 313; Treatment total *n* = 171; Control total *n* = 142.

The multilevel analysis revealed no sustained intervention effect at the second year follow-up, $\beta$ = -2.31, *p* = .076. To assess how this same subsample of students performed at the intervention year follow-up (i.e., whether there was an initial intervention effect to sustain), the same analysis was conducted with this subsample using intervention year follow-up OAKS NCE scores. Similar to the second year follow-up results, no initial intervention effect was seen for this subsample, $\beta$ = 2.00, *p* = .084. Thus this subsample of students—the only available subsample with which to examine sustained effects of the intervention—is not a sufficient subsample to address the question of sustained effects.

# Appendix L
# Fixed and Random Effects and Intraclass Correlations

# Fixed and Random Effects and Intraclass Correlations (ICCs)

**Exhibit L.1**
**Final Model for Estimating Fixed and Random Effects of Xtreme Reading Impact on Spring GRADE Total NCE Scores (Overall GRADE Sample)**

| Fixed Effects | | | | | | |
|---|---|---|---|---|---|---|
| **Level** | **Effect** | **Impact (β)** | ***SE*** | ***DF*** | ***t*** | ***p*** |
| School | Intercept | 30.27 | 0.79 | 11 | 38.24 | .000 |
| Student | Xtreme Reading | 3.34 | 0.60 | 1566 | 5.60 | .000 |
| | Fall GRADE Total Score | 0.71 | 0.02 | 1566 | 31.92 | .000 |
| | $D_{GRADE}$[a] | 20.02 | 1.14 | 1566 | 17.55 | .000 |
| | Black | -2.91 | 0.79 | 1566 | -3.68 | .000 |
| | Hispanic | -0.70 | 0.75 | 1566 | -0.94 | .349 |
| | ELL | -3.51 | 0.74 | 1566 | -4.73 | .000 |
| | Cohort 2 | 0.32 | 0.94 | 1566 | 0.34 | .730 |
| | Cohort 3 | 2.62 | 0.81 | 1566 | 3.23 | .002 |
| | Cohort 4 | 1.34 | 0.90 | 1566 | 1.51 | .131 |
| | Grade 8 | 1.49 | 0.86 | 1566 | 1.74 | .081 |
| | Grade 9 | -1.68 | 1.41 | 1566 | -1.19 | .236 |
| | Grade 10 | 2.42 | 1.48 | 1566 | 1.64 | .100 |

| Random Effects | | |
|---|---|---|
| **Level** | **Variance Components** | **Variance** |
| School | Level 2 | 5.91 |
| Student | Level 1 | 137.57 |

| Random Effects (From Unconditional Model)[b] | | | |
|---|---|---|---|
| **Level** | **Variance Components** | **Variance** | **ICC** |
| School | Level 2 | 11.76 | 0.043 |
| Student | Level 1 | 262.36 | |

[a]$D_{GRADE}$ = indicator for missing pretest data (0 = non-missing, 1 = missing). [b]The unconditional model is a 2-level model with students (Level 1) nested in schools (Level 2), and only an intercept term on the right-hand side of the model.

## Exhibit L.2
## Final Model for Estimating Fixed and Random Effects of Xtreme Reading Impact on Spring GRADE Total NCE Scores (Middle School GRADE Sample)

**Fixed Effects**

| Level | Effect | Impact (β) | *SE* | *DF* | *t* | *p* |
|---|---|---|---|---|---|---|
| School | Intercept | 29.68 | 1.11 | 9 | 26.83 | .000 |
| Student | Xtreme Reading | 4.39 | 0.69 | 811 | 6.31 | .000 |
| | Fall GRADE Total Score | 0.71 | 0.03 | 811 | 26.67 | .000 |
| | $D_{GRADE}$[a] | 16.41 | 1.51 | 811 | 10.86 | .000 |
| | Black | -2.94 | 0.95 | 811 | -3.09 | .003 |
| | Hispanic | -1.12 | 0.84 | 811 | -1.33 | .183 |
| | ELL | -2.59 | 0.80 | 811 | -3.21 | .002 |
| | Cohort 2 | -1.63 | 1.16 | 811 | -1.41 | .159 |
| | Cohort 3 | -1.03 | 1.12 | 811 | -0.92 | .358 |
| | Cohort 4 | -2.86 | 1.17 | 811 | -2.44 | .015 |
| | Grade 8 | 1.31 | 0.74 | 811 | 1.77 | .077 |

**Random Effects**

| Level | Variance Components | Variance |
|---|---|---|
| School | Level 2 | 9.70 |
| Student | Level 1 | 97.47 |

**Random Effects (From Unconditional Model)[b]**

| Level | Variance Components | Variance | ICC |
|---|---|---|---|
| School | Level 2 | 15.25 | 0.068 |
| Student | Level 1 | 210.65 | |

[a]$D_{GRADE}$ = indicator for missing pretest data (0 = non-missing, 1 = missing). [b]The unconditional model is a 2-level model with students (Level 1) nested in schools (Level 2), and only an intercept term on the right-hand side of the model.

## Exhibit L.3
## Final Model for Estimating Fixed and Random Effects of Xtreme Reading Impact on Spring GRADE Total NCE Scores (High School GRADE Sample)

**Fixed Effects**

| Level | Effect | Impact (β) | *SE* | *DF* | *t* | *p* |
|---|---|---|---|---|---|---|
| School | Intercept | 30.41 | 0.92 | 4 | 33.22 | .000 |
| Student | Xtreme Reading | 2.07 | 0.98 | 746 | 2.12 | .035 |
| | Fall GRADE Total Score | 0.71 | 0.04 | 746 | 19.76 | .000 |
| | $D_{GRADE}$[a] | 22.20 | 1.73 | 746 | 12.83 | .000 |
| | Black | -2.63 | 1.25 | 746 | -2.11 | .035 |
| | Hispanic | 0.21 | 1.39 | 746 | 0.17 | .868 |
| | ELL | -4.72 | 1.35 | 746 | -3.49 | .001 |
| | Cohort 2 | 2.50 | 1.46 | 746 | 1.71 | .087 |
| | Cohort 3 | 5.04 | 1.23 | 746 | 4.10 | .000 |
| | Cohort 4 | 4.45 | 1.39 | 746 | 3.21 | .002 |
| | Grade 10 | 4.08 | 1.00 | 746 | 4.07 | .000 |

**Random Effects**

| Level | Variance Components | Variance |
|---|---|---|
| School | Level 2 | 2.61 |
| Student | Level 1 | 175.94 |

**Random Effects (From Unconditional Model)[b]**

| Level | Variance Components | Variance | ICC |
|---|---|---|---|
| School | Level 2 | 8.22 | 0.025 |
| Student | Level 1 | 317.76 | |

[a]$D_{GRADE}$ = indicator for missing pretest data (0 = non-missing, 1 = missing).[b]The unconditional model is a 2-level model with students (Level 1) nested in schools (Level 2), and only an intercept term on the right-hand side of the model.

## Exhibit L.4
## Final Model for Estimating Fixed and Random Effects of Xtreme Reading Impact on Spring OAKS Total Reading Scores (Overall OAKS Sample)

**Fixed Effects**

| Level | Effect | Impact (β) | *SE* | *DF* | *t* | *p* |
|---|---|---|---|---|---|---|
| School | Intercept | 32.42 | 0.86 | 11 | 37.83 | .000 |
| Student | Xtreme Reading | 1.26 | 0.58 | 1454 | 2.16 | .031 |
| | OAKS Pretest Score | 0.30 | 0.03 | 1454 | 9.76 | .000 |
| | $D_{OAKS}$[a] | 6.93 | 1.35 | 1454 | 5.14 | .000 |
| | Fall GRADE Total Score | 0.31 | 0.02 | 1454 | 12.55 | .000 |
| | $D_{GRADE}$[b] | 8.19 | 1.08 | 1454 | 7.54 | .000 |
| | Black | -2.75 | 0.77 | 1454 | -3.58 | .001 |
| | Hispanic | -0.52 | 0.74 | 1454 | -0.71 | .478 |
| | ELL | -1.94 | 0.74 | 1454 | -2.62 | .009 |
| | Cohort 2 | -3.63 | 0.91 | 1454 | -3.98 | .000 |
| | Cohort 3 | -6.59 | 0.86 | 1454 | -7.64 | .000 |
| | Cohort 4 | -5.49 | 0.97 | 1454 | -5.66 | .000 |
| | Grade 8 | -3.51 | 0.75 | 1454 | -4.68 | .000 |
| | Grade 10 | 0.42 | 1.38 | 1454 | 0.31 | .757 |

**Random Effects**

| Level | Variance Components | Variance |
|---|---|---|
| School | Level 2 | 7.54 |
| Student | Level 1 | 123.59 |

**Random Effects (From Unconditional Model)** [c]

| Level | Variance Components | Variance | ICC |
|---|---|---|---|
| School | Level 2 | 8.86 | 0.048 |
| Student | Level 1 | 175.22 | |

[a]$D_{OAKS}$ = indicator for missing OAKS pretest data (0 = non-missing, 1 = missing). [b]$D_{GRADE}$ = indicator for missing Fall GRADE NCE data (0 = non-missing, 1 = missing). [c]The unconditional model is a 2-level model with students (Level 1) nested in schools (Level 2), and only an intercept term on the right-hand side of the model.

## Exhibit L.5
## Final Model for Estimating Fixed and Random Effects of Xtreme Reading Impact on Spring OAKS Total Reading Scores (Middle School OAKS Sample)

**Fixed Effects**

| Level | Effect | Impact (β) | *SE* | *DF* | *t* | *p* |
|---|---|---|---|---|---|---|
| School | Intercept | 30.98 | 1.32 | 9 | 23.39 | .000 |
| Student | Xtreme Reading | 1.69 | 0.69 | 941 | 2.45 | .014 |
| | OAKS Pretest Score | 0.30 | 0.03 | 941 | 8.71 | .000 |
| | $D_{OAKS}$[a] | 7.36 | 1.95 | 941 | 3.78 | .000 |
| | Fall GRADE Total Score | 0.32 | 0.03 | 941 | 10.57 | .000 |
| | $D_{GRADE}$[b] | 6.58 | 1.37 | 941 | 4.79 | .000 |
| | Black | -1.53 | 0.92 | 941 | -1.64 | .102 |
| | Hispanic | 0.02 | 0.84 | 941 | 0.02 | .981 |
| | ELL | -1.29 | 0.80 | 941 | -1.60 | .108 |
| | Cohort 2 | -0.79 | 1.13 | 941 | -0.70 | .487 |
| | Cohort 3 | -6.19 | 1.14 | 941 | -5.41 | .000 |
| | Cohort 4 | -4.90 | 1.20 | 941 | -4.08 | .000 |
| | Grade 8 | -3.19 | 0.74 | 941 | -4.35 | .000 |

**Random Effects**

| Level | Variance Components | Variance |
|---|---|---|
| School | Level 2 | 14.86 |
| Student | Level 1 | 110.87 |

**Random Effects (From Unconditional Model)[c]**

| Level | Variance Components | Variance | ICC |
|---|---|---|---|
| School | Level 2 | 19.08 | 0.105 |
| Student | Level 1 | 162.37 | |

[a]$D_{OAKS}$ = indicator for missing OAKS pretest data (0 = non-missing, 1 = missing). [b]$D_{GRADE}$ = indicator for missing Fall GRADE NCE data (0 = non-missing, 1 = missing). [c]The unconditional model is a 2-level model with students (Level 1) nested in schools (Level 2), and only an intercept term on the right-hand side of the model.

## Exhibit L.6
## Final Model for Estimating Fixed and Random Effects of Xtreme Reading Impact on Spring OAKS Total Reading Scores (High School OAKS Sample)

**Fixed Effects**

| Level | Effect | Impact (β) | S.E. | DF | *t* | *p* |
|---|---|---|---|---|---|---|
| School | Intercept | 35.03 | 0.53 | 4 | 66.45 | .000 |
| Student | Xtreme Reading | 0.29 | 1.06 | 502 | 0.28 | .784 |
| | OAKS Pretest Score | 0.24 | 0.05 | 502 | 4.61 | .000 |
| | $D_{OAKS}$[a] | 6.17 | 2.15 | 502 | 2.87 | .005 |
| | Fall GRADE Total Score | 0.30 | 0.04 | 502 | 6.73 | .000 |
| | $D_{GRADE}$[b] | 8.48 | 1.81 | 502 | 4.68 | .000 |
| | Black | -4.03 | 1.22 | 502 | -3.29 | .001 |
| | Hispanic | -1.15 | 1.48 | 502 | -0.78 | .438 |
| | ELL | -4.13 | 1.68 | 502 | -2.45 | .015 |
| | Cohort 2 | -7.07 | 1.58 | 502 | -4.48 | .000 |
| | Cohort 3 | -6.83 | 1.44 | 502 | -4.75 | .000 |
| | Cohort 4 | -6.78 | 1.73 | 502 | -3.92 | .000 |

**Random Effects**

| Level | Variance Components | Variance |
|---|---|---|
| School | Level 2 | 0.01 |
| Student | Level 1 | 141.99 |

**Random Effects (From Unconditional Model)[c]**

| Level | Variance Components | Variance | ICC |
|---|---|---|---|
| School | Level 2 | 1.66 | 0.009 |
| Student | Level 1 | 187.02 | |

[a] $D_{OAKS}$ = indicator for missing OAKS pretest data (0 = non-missing, 1 = missing).
[b] $D_{GRADE}$ = indicator for missing Fall GRADE NCE data (0 = non-missing, 1 = missing).
[c] The unconditional model is a 2-level model with students (Level 1) nested in schools (Level 2), and only an intercept term on the right-hand side of the model.