

# Chicago Public Schools

# Striving Readers Initiative

## Year Two Evaluation Report

---

January 2009

Principal Investigator:

**Alan J. Simon, Ph.D., Senior Vice President**

Project Director/Primary Author:

**Jonathan Tunik, M.P.A., Senior Associate**

Design Consultation, Data Analysis and Data Management and General Support:

**Joy Zacharia, Senior Research Associate**

**Tania Tasse, Senior Research Associate**

**Julia Alemany, Research Associate**

**Jing Zhu, Research Associate**

**Rebecca Swann, Research Analyst**

**Allison Fields, Research Assistant**

Metis Associates

90 Broad Street, Suite 1200

NY, NY 10004

212-425-8833

212-480-2176 – fax

[www.metisassociates.com](http://www.metisassociates.com)

## Table of Contents

---

Table of Contents .....	i
I. Executive Summary of Findings: Implementation and Impact .....	iii
Variations by Component and by School in Fidelity of Implementation: .....	iii
Whole-school, Targeted and Intensive Interventions, Years 1 and 2 .....	iii
Year 1 Implementation Study .....	iii
Year 2 Implementation Study .....	iv
Impacts on Teacher Practices and Classroom Instructional Environment.....	vi
Whole-school Intervention.....	vi
Targeted and Intensive Interventions.....	vi
Impacts on Students .....	vi
Overall Program Impacts .....	vii
Targeted and Intensive Interventions.....	vii
Conclusions.....	vii
II. Introduction and Study Background .....	1
Context for the Study .....	1
Theoretical Rationale for and Description of the Intervention Models .....	2
Administrative Structure.....	2
Targeted and Intensive Interventions.....	3
Whole-School (Blended) Intervention .....	4
Targeted Schools, Grades, and Students.....	6
Logic Model.....	9
Brief Overview of Key Evaluation Design Features .....	11
Research Questions .....	11
III. Evaluation of the Implementation of the Whole School, Targeted, and Intensive Interventions: Years 1 and 2.....	14
Study Design.....	14
Research Questions .....	14
Data Sources .....	14
Year 1 Implementation Study .....	17
Intervention as Implemented.....	17
Implications for Impact Analyses .....	21
Year 2 Implementation Study .....	22
Intervention as Implemented.....	22
Implications for Impact Analyses .....	29
Comparison of Year 1 and Year 2 Implementation .....	29
IV. Evaluation of the Impact of the Targeted and Intensive Interventions: Years 1 and 2 .....	32
Study Design.....	32
Research Questions .....	32
Sampling Plan .....	32
Sample Size and Power.....	34
Data Collection Plan .....	35
Summary of Analytic Approach to the Impact Analysis .....	36
Description of First- and Second-Year Samples.....	37
Numbers and Basic Characteristics of Schools, Teachers, Classrooms, and Students.....	37

Impacts on Students at the End of Two Years .....	43
Additional Analyses .....	44
V. Evaluation of the School-Wide Impact of the Chicago Striving Readers Initiative in Years 1 and 2.....	49
Study Design.....	49
Research Questions .....	49
Sampling Plan .....	49
Data Collection Plan .....	50
Summary of Analytic Approach to the Impact Analysis .....	51
Description of the First- and Second-Year Samples.....	52
Number and Basic Characteristics of Schools, Teachers, Classrooms, and Students in First-Year Sample.....	52
Impacts on Teacher Practices and Classroom Instructional Environment at the End of Two Years .....	52
Use of High-Quality, High-Interest Materials .....	53
Use of Core Comprehension Strategies .....	55
Use of Vocabulary Instruction.....	58
Use of Small-Group Instruction.....	59
Targeted Intervention: Presence of the LIT .....	60
Targeted Intervention: Small-Group Instruction Provided by the LIT .....	61
Intensive Intervention: After-School Literacy Program Provided by the LIT.....	62
Discussion and Conclusions .....	63
Impacts on Students .....	63
Additional Analyses.....	64
VI. Summary and Conclusions .....	68
References.....	72
APPENDICES .....	73

# I. Executive Summary of Findings: Implementation and Impact

This report describes the evaluation of the second year of implementation of the five-year Chicago Striving Readers program. The evaluation involved 31 treatment schools (16 from Cohort 1 and 15 from Cohort 2) and 32 control schools (16 per cohort).

## **Variations by Component and by School in Fidelity of Implementation: Whole-school, Targeted and Intensive Interventions, Years 1 and 2**

For first and second year implementation results, it was in many cases difficult to assign evidence about implementation to specific intervention models (whole-school, targeted or intensive) because the phrasing of the data collection instruments often was not explicit about the context. Thus, most of the evidence reflects fidelity of implementation of the overall program, rather than any particular intervention model. Findings that could be explicitly linked to the targeted and intensive interventions are presented wherever possible. Findings from Year 1 are summarized from the Year 1 Implementation Report.

### *Year 1 Implementation Study*

#### *Overall Program Implementation*

Although there were some delays in integrating some of the program-related **technology**, most staff reported that technology was at least *somewhat integrated* by spring 2007 and the majority of teachers also reported spending at least *some* instructional time using computers and technology. Half of Striving Readers teachers surveyed reported that they used listening centers at least once a week; use of media centers increased during the course of the year, with 60% of teachers observed during the fall of 2006 using them, and 85% of those observed in spring 2007.

Roughly half of all teachers were observed using a **whole-part-whole instructional framework**, although these teachers only spent a “moderate” amount of time doing so. A majority of teachers in all classrooms and content areas also reported that they regularly covered at least some of the **comprehension strategies**. The majority of teachers reported **adjusting instructional practices based on diagnostic and assessment results**, and most indicated that they worked with their schools’ Lead Literacy Teachers (LLTs) to use assessment data for instructional planning. Insufficient data were available to determine whether the **gradual release model** was being used, although important components of the model were observed fairly regularly.

However, a number of program components were not yet fully in place—and in some cases, not in place at all—during Year 1, and others had a slow start early in the year. While Literacy Intervention Teachers (LITs) were trained on **Partner Reading in the Content Area, Too (PRC2)** and reported being prepared to immediately introduce this instructional method, the materials were not available during Year 1. Similarly, **text sets**, which were not acquired until near the end of the year, were only used approximately one third of the time. Finally, although two fifths of teachers surveyed felt that the program was *very effective* in helping them develop the use of partner reading for vocabulary development and felt that it was at least *moderately*

*effective*, only about one fifth to one third of all observed language arts classes included small-group activities focusing on **vocabulary**.

Part of the reason for some of these delays in implementation is that it takes time for teachers to become comfortable using particular instructional strategies and methods. Initial delays may also have been due in part to delays in hiring staff: although LITs were on board when the program started, they were not hired early enough to complete their own orientation training before the program started, as originally planned. This may have made it more difficult for them to support teachers' implementation efforts early in the year—although they reported that their collaborations with teachers had increased by mid year, as discussed below.

### ***Targeted and Intensive Interventions***

**Collaboration of teachers and LITs** was fostered through initial professional development activities. By winter 2007, all LITs reported actively collaborating with teachers. Teachers and LITs also collaborated through school Literacy Leadership Teams and grade-level teams, which were meeting in 63% and 67%, respectively, of the schools by spring 2007. The targeted intervention model included **direct, targeted instruction of Tier 2 students during 20-minute periods of small-group instruction** by LITs, which took place in 46% to 59% of the observed literacy blocks, although it is likely that it was also occurring at times other than when the observations took place. As part of the intensive intervention, **240 additional minutes of supported literacy instruction** was systematically provided to Tier 3 students through the AMP after-school program. However, AMP attendance, averaging 81%, might not have been high enough to ensure that all Tier 3 students received the full intervention. AMP classes took place in **small-group settings** where there was an average of one teacher for every 10 students, exceeding the goal of a student-teacher ratio of 15:1. **Guided fluency practice** was provided as part of the AMP program. Although AMP was implemented in all after-school programs, however, its impact may have been limited somewhat by the fact that, as LITs noted, some topics were of limited relevance to students and others became tedious for them.

### ***Year 2 Implementation Study***

Data sources that were used in Year 2 to provide evidence about key program features—including staff interviews, teacher surveys, observations and program schedules and attendance records—a series of rubrics were created that were used to generate scores representing level of program fidelity within the classroom and professional development models.<sup>1</sup> On a 10-point scale, on average, all schools in both cohorts were implementing the Striving Readers program at a medium level of overall fidelity (with an average score of 6.6) during the 2007–2008 program year, with no overall difference between cohorts. Among the major program components, the highest mean ratings were obtained for the intensive intervention, for which fidelity ratings averaged 7.8. Further details about results for the three program models follow.

### ***Whole-school Implementation***

Within this component, the highest mean fidelity ratings were obtained for use of the whole-part-whole instructional model and the gradual release model, both of which received mean fidelity

---

<sup>1</sup> Because of the transition to a new evaluator during the second program year, these scores could not be calculated for Year 1.

scores across all schools at the “high implementing” level (averaging 8.9 and 8.5, respectively). Only one Cohort 1 school and three Cohort 2 schools fell slightly short of the “high implementing” level of fidelity on the gradual release model. Implementation appeared to be the most problematic for anchoring instruction in comprehension, for which ratings averaged 5.2.

Compared with Year 1, the use of PRC2 showed the most dramatic improvement in Year 2. However, program staff’s assessment of program implementation indicated that while PRC2 may have been implemented regularly in Year 2, it was often not implemented correctly. Similarly, teachers’ self-reports indicated an increase in use of the whole-part-whole classroom instruction model, although it is not clear whether they were implementing this model as designed.

### ***Targeted and Intensive Interventions***

The most successful specific program component, according to fidelity scale results, was the creation of small-group settings during the after-school program: all schools in both cohorts met, or at least came very close to, the required 15:1 student-to-teacher ratio, with mean fidelity scores of 9.5 among Cohort 1 schools and 9.8 in Cohort 2. Nevertheless, there was still considerable variability by school, several of which did not reach the “high implementing” level of fidelity on this sub-component.

One area that was a challenge to program implementation in both years was increased instructional time in the AMP program for Tier 3 students, which was hindered in part because of irregular student attendance, and in part due to program schedules at some schools. In Year 2, irregular attendance resulted in Cohort 2 schools receiving the lowest rating for this sub-component of any of the fidelity scales.

As would be expected, implementation varied considerably among schools, although it did more so for some program components than others. Implementation varied most significantly on ratings of collaboration between the LIT and classroom teachers, and on increased instructional time for the intensive intervention, both of which ranged from perfect or near perfect scores to the low end of the “low implementing” level of fidelity. Substantial variations among schools were also observed for direct vocabulary instruction and implementation of whole-part-whole instructional structures.

In efforts to improve program fidelity, a substantial number of respondents to school leader interviews conducted in Year 2 cited the importance of the professional development program, particularly for helping them integrate literacy into other content areas. Respondents to the spring 2008 teacher survey also expressed a desire for more training and support. However, the professional development component was rated at a “medium implementing” level of fidelity for most schools (5.5 overall), implying that principals, LITs, and/or teachers had low average attendance rates at least some of the professional development sessions, with lower attendance rates among Cohort 1 schools.

On all of the above fidelity scores, it should be cautioned that the extent to which the each scale fully reflects all aspects of the program model varies from component to component. For the Year 3 evaluation, alignment between fidelity scale results and district staff’s observations of

program fidelity is being studied in order to inform the improvement of data collection instruments on which the scales are based.

## Impacts on Teacher Practices and Classroom Instructional Environment

Data were available from spring 2008 classroom observations that provided evidence about impact for several key program components relevant to the whole-school/blended intervention model, as well as some that are specific to the targeted intervention.<sup>2</sup> Any differences that are observed between treatment and control schools in the availability and use of resources or in changes in the way in which the resources and instructional methods are applied can be taken as supporting evidence of the impact of the program on classroom practices.

### *Whole-school Intervention*

Observed **use of high-quality, high-interest materials** was higher among treatment classrooms—more treatment classrooms had appropriate media centers and the majority of treatment classrooms had listening centers; however, treatment and control classrooms were similar in terms of the availability and quality of libraries accessible to students. **Comprehension strategies** were used more frequently in treatment classrooms in both years, but both groups showed a marked increase in Year 2. *Summarizing* was the most frequently used comprehension strategy, and treatment classes were observed conducting instructional activities using *visualizing* (in both cohorts) and *metacognition* (in Cohort 2) more frequently than control classrooms. However, more control classes were observed using *inferring* (both Cohorts) and *predicting* and *metacognition* (Cohort 1). Substantially more treatment classrooms from both cohorts demonstrated direct **vocabulary instruction**. In Year 2, a substantially greater proportion of treatment classrooms were observed using **small groups** and/or pairs. Nearly all Cohort 1 and 2 treatment classrooms grouped students into small groups and/or pairs for literacy instruction, compared to 53% of Cohort 1 control and 73% of the Cohort 2 control classrooms.

### *Targeted and Intensive Interventions*

Data from spring 2008 classroom observations provided some evidence specific to the targeted intervention. During the observations, the role of the LIT was found to be essentially unique to Striving Readers schools. Far more treatment classrooms than control classrooms had an **adult other than the teacher present** (the LIT in treatment classrooms versus any other adult in control classrooms) during the literacy block. This was true for Cohort 1 classrooms in both years, and in both cohorts in Year 2. The proportion of treatment classes where an LIT was present also increased substantially from Year 1 to Year 2. In addition, **small-group instruction led by the LIT** was essentially unique to the treatment classes. In Year 2, an adult was observed providing small-group instruction in only 2% of the Cohort 1 and Cohort 2 control school classes, compared with 75% of Cohort 1 and 78% of Cohort 2 treatment classes.

## Impacts on Students

Hierarchical linear modeling (HLM) was used to estimate the impacts of the Striving Readers program on students' reading performance, as measured by their reading scores on the Illinois

---

<sup>2</sup> Most of the AMP classes had already ended by the time spring observations were conducted, resulting in insufficient data to provide evidence about the intensive intervention.

Student Achievement Test. Cross-sectional impact analyses focused on the intent-to-treat populations, and used two-level HLM models in order to account for the clustering of students in schools. Analyses were designed to assess impacts on three groups of students: the overall program impacts on all students, the combined impact of the whole-school and targeted interventions on Tier 2 students, and the impacts of the whole-school, targeted and intensive interventions on Tier 3 students. Analyses of overall program impact examined all participating grades, while the Tier 2 and Tier 3 impact analyses focused on the 6th grade.

### ***Overall Program Impacts***

Analyses revealed several contributing factors which influence reading performance: individual baseline test scores in both reading and mathematics were positive predictors of spring 2008 reading performance, and students in smaller schools and students in Cohort 1 schools were more likely to score higher than their peers. In contrast, being African American, a male, or a student in special education were significant negative predictors of students' reading performance. Because of the vertical scaling of the ISAT and the selection process for assigning students to tiers, respectively, students in grade 8 and students in Tier 1 were more likely to score higher than their peers after holding all other covariates constant. While analyses did not reveal any significant overall impact of the Striving Readers Initiative on students' reading performance, they did reveal that the program has been more effective with students in 6th- and 7th-grade than with 8th-grade students.

### ***Targeted and Intensive Interventions***

Again using HLM to assess impact on spring 2008 ISAT scores for the intent-to-treat population, prior academic achievement was again found to be a positive predictor of spring 2008 performance, and schools with higher proportions of students reading at or above grade level also tended to produce higher individual reading performance on the posttest, even among students whose individual baseline scores were below grade level. In contrast, students at schools with larger proportions of low-income students were more likely to score lower on the spring 2008 ISAT, regardless of their own socioeconomic status. However, there were no detectable overall impacts of the program on 6th grade students in Tiers 2 or 3.

### **Conclusions**

There are several possible reasons why an impact on student performance was not detected. The initiative is still in its early stages, and so implementation is still not fully aligned with the model. Results of the intent-to-treat model speak to whether the program is working as implemented, but they do not necessarily reflect whether the intended *model* would work if implemented correctly. The findings can also be confounded by the possibility that the control schools are aware of the Striving Readers program and may decide to adopt some of the initiatives, methods, or resources on their own. Nevertheless, it is clear that further efforts to strengthen program implementation could increase the likelihood of affecting more positive outcomes for students, as well as improve the chances of detecting the initiative's potential to impact student achievement.



## II. Introduction and Study Background

---

### Context for the Study

In 2005, the U.S. Department of Education's Office of Elementary and Secondary Education issued a request for proposals (RFP) for programs to improve adolescent literacy. The Department's stated goal was to improve the performance of struggling adolescent readers and help build a strong scientific research base around specific strategies to help struggling readers. In response to the RFP, the Chicago Public Schools (CPS) proposed the Chicago Striving Readers program. This five-year program aims to transform teaching and learning through a seamless, aligned approach to literacy instruction for grades 6–8 across 31 schools.

The framework of the Chicago Striving Readers program was designed to transform the teaching and learning of middle-grade comprehension-focused literacy strategies and to drive long-term, systemic improvement in adolescent literacy in the district through a research-based model of prevention and intervention. The Chicago Striving Readers framework, which is based on an analysis of data and research reviews, encompasses optimal instructional strategies and infrastructural support services to improve the reading achievement of CPS middle-grade students. The framework provides a seamless and unifying, yet flexible, system that redesigns how reading is taught in grades 6–8.

The Chicago Striving Readers program uses a tiered approach to working with adolescent readers. At the start of the school year, students are tiered into three groups: good readers (Tier 1), struggling readers who could reach grade level with focused support in the classroom (Tier 2); and struggling readers who require long-term intensive support and customized instruction (Tier 3). All three tiers receive the benefits of a whole-school blended intervention. Tiers 2 and 3 benefit from both whole-school and targeted interventions, and Tier 3 students benefit from the whole-school, targeted, and intensive interventions.

The Chicago Striving Readers program has seven key components, including three intervention models:

1. A whole-school, blended intervention model consisting of reading comprehension instruction for all students (Tiers 1, 2, and 3) in grades 6–8;
2. A targeted intervention model consisting of reading comprehension instruction for struggling readers (Tiers 2 and 3) in grades 6–8;
3. An intensive intervention model consisting of reading comprehension instruction for struggling readers (Tier 3) primarily in grade 6<sup>3</sup>;
4. Frequent, purposeful assessment and adjustment of instruction with screening, diagnostic, and progress-monitoring tools;
5. Data-driven instruction structured through a team-based system of leadership and support;
6. High-quality, high-interest materials that are integrated with engaging technology and audio resources; and

---

<sup>3</sup> The intensive intervention model is also provided to Tier 3 students in Grades 7 and 8 by classroom teachers to the extent possible; however, grant funds were only sufficient to directly support this intervention for 6th graders.

## 7. Integrated, progressive, and high-quality professional development.

The literacy achievement gaps of middle-grade students in the Chicago Striving Readers program are an outgrowth of complex socioeconomic and educational factors endemic to large urban areas like Chicago. Research in adolescent literacy has established that the most common problem of adolescent struggling readers is that they are not able to comprehend what they read (Biancarosa & Snow, 2004). In the 16 schools where the Striving Readers Initiative was implemented during the first year (Cohort 1), more than half (53%) of the students do not meet reading standards at the end of 5th grade. Although this deficit is reduced to 41% by the end of 8th grade, it has already taken a large toll on student achievement in other subject areas. The vast majority of schools in the Chicago Striving Readers program are located in communities that are economically disadvantaged and racially segregated. An average of 90% of students in these schools receive free or reduced-price lunches, and the No Child Left Behind (NCLB) Title I Poverty Indices range from 49% to 68%. Geographic isolation by race and ethnicity compounds the pedagogical challenges of educating students with learning disabilities and English language learners (ELLs) who come from homes where English is not always spoken.

### **Theoretical Rationale for and Description of the Intervention Models**

No single explanation accounts for why some students struggle with reading after the 5th grade while others do not. Although struggling adolescent readers are often characterized by terms such as at-risk, unmotivated, disenchanted, or generally unsuccessful in school literacy tasks, research suggests that these descriptors are secondary consequences of underlying problems, not the primary causes (Peterson *et al.*, 2000; Moats, 2001). Sociocultural, motivational, and linguistic factors may be involved to varying degrees, but most of the research focuses on a cognitive basis or “deprivation approach” as the major underlying problem. This approach assumes that students must master a stable set of tasks or milestones to qualify as developmentally competent readers. Below-average performance on these tasks indicates that students have not developed the requisite skills necessary for reading competency at a particular grade or in a particular set of tasks. The assumption is that below-average performance indicates deficits in varying combinations of word recognition and decoding skills; language processing ability at the word, sentence, or conversation levels; vocabulary; background knowledge; awareness of one’s own comprehension processes (metacognition); and comprehension/study strategies (Moore, Alvermann, & Hinchman, 2000).

The intervention component of the Chicago Striving Readers program is based on the deprivation approach—CPS identifies students who are struggling with reading, tracks the nature and state of their deficits on an ongoing basis, and provides intensive and targeted support within school-wide language arts and subject-area classes and through an extended-day class.

### ***Administrative Structure***

The Chicago Striving Readers Initiative is managed by a strong team of leaders in the field of literacy instruction. The roles of the key players on this administrative team are summarized below.

**Project Director.** Ms. Elizabeth Cárdenas-Lopez, who had served as one of the Striving Readers Project Coordinators during the first program year, was appointed as Project Director in June 2008. The Project Director is responsible for overseeing day-to-day project operation, maintaining communications and meeting regularly with staff and partners, ensuring fiscal integrity and adherence to grant requirements, monitoring and ensuring the quality of professional development, managing data collection and dissemination and interfacing with the evaluator, and coordinating general scheduling. Ms. Cárdenas-Lopez also serves a broader role building departmental capacity and financial sustainability by working with Office of Literacy to develop middle school curriculum, selecting classroom materials, creating diagnostic assessments, and managing Office of Literacy meetings.

**Senior Literacy Advisor.** Dr. Donna Ogle, Professor of Reading and Language at National-Louis University, holds the position of Senior Literacy Advisor, serving on both the Striving Readers advisory board and on the district team. Dr. Ogle is a specialist and an innovator in instructional strategies that support increased literacy and instructional change in schools. Her responsibilities include providing both group and individual leadership training for Striving Readers Coordinators and for school principals and Lead Literacy Teachers. In addition, Dr. Ogle provides a critical, ongoing advisory role during individual meetings with the Project Director. She has also taken on a degree of responsibility for supporting the LITs' efforts in coordinating literacy course-work at the schools.

**Project Coordinators.** The four Striving Readers School Coordinators—each of whom has direct responsibility for seven to eight schools—provide the schools with classroom instructional guidance and support and facilitate instructional planning. This includes conducting school-based professional development and one-on-one technical assistance for both teachers and LITs, and assisting LITs with observing and modeling lessons for teachers, providing lesson planning support. While the Coordinators work primarily with language arts classrooms and teachers, they also provide support to other subject area teachers through the departmental and literacy team meetings. Coordinators also work directly with school principals, although the extent of this aspect of their role varies depending on the principal's needs. Work with principals might include one-on-one coaching and support, planning of school-wide Professional Development and conferences, and discussion of specific classroom needs. While the majority of their time is spent on-site at the schools, the Coordinators also participate in meetings with other program leaders at the district office, in addition to their interactions with the Senior Literacy Advisor while they are receiving training.

### ***Targeted and Intensive Interventions***

To meet the needs of struggling readers, the Chicago Striving Readers program features a two-pronged approach: (1) small group, focused instruction for Tier 2 and Tier 3 students that is blended within language arts and subject area classes in the school-wide model (targeted intervention); and (2) intensive, strategic, targeted instruction for approximately 15–20 Tier 3 students in grades 6–8 that occurs after the regular school day (intensive intervention).

The targeted intervention within the school-wide model starts as soon as possible after the students have been tiered and continues for the duration of the school year. It involves differentiated instruction and scaffolding within regular classroom settings as well as increased individualized in-class support during the crucial 6th-grade year, when students are laying the

foundation for middle grade and future academic success. These activities take place during the regular ELA class, while Tier I students continue to receive instruction from the classroom teacher according to the Whole School Intervention model described below.

The intensive intervention through the extended-day approach also begins as soon as possible (October) after Tier 3 students are identified, and continues through May. Literacy Intervention Teachers (LITs) conduct the after-school classes for 6th-grade students, and classroom teachers conduct the classes for students in grades 7 and 8. Intervention at this level features one hour of customized instruction for small homogeneous groups on four days each week. This provides: (1) an additional 240 minutes each week of direct and supported instruction beyond the intervention that occurs during the regular school day; (2) small group instruction in groups of three to four students per teacher (within a maximum class size of 15 students to one teacher); (3) more frequent assessment and adjustment of instruction; and (4) highly motivating reading materials integrated with technology and audio.

The LIT is central to the targeted intervention component of the Striving Readers program. These teachers divide their time between 6th-grade classrooms to assist in providing differentiated literacy instruction. The classroom teacher is responsible for providing the overall literacy instruction in the classroom, assigning grades, and collaborating with the LITs to create weekly lesson and intervention plans. The LITs and the 6th-grade classroom teachers meet regularly to prioritize and coordinate instruction (e.g., skill review, assessment, explicit teaching, responsibilities, student groupings), and in small schools, LITs also meet with 7th- and 8th-grade classroom teachers. The focus of the LITs on targeted, differentiated instruction to meet the needs of all students within the regular classroom is of particular benefit to Tier 2 and Tier 3 students who are struggling, including ELL students and students with learning disabilities. The identification of specific literacy challenges of Tier 2 and Tier 3 students enables LITs and teachers to focus literacy instruction on the reading comprehension and vocabulary needs of these students. The use of differentiated materials and monitoring of the students' progress is critical to the targeted intervention.

### ***Whole-School (Blended) Intervention***

The whole-school intervention occurs within both language arts and subject-area classrooms. In language arts classes, direct explicit instruction in comprehension, fluency, vocabulary, and word identification occurs within a research recommended 90-minute instructional model. This model facilitates the gradual release of responsibility from teacher to students within a whole group/small group/whole group (whole-part-whole) configuration. Research by Atwell (1998) informs the reader/writer workshop approach for small groups. Students start with 10–20 minutes of whole-group instruction; move to 60 minutes of practice and application in small groups; and conclude with 10–20 minutes of whole-group sharing. During the small-group period, students rotate, as assigned, among three 20-minute workshop activities.

The instructional organization of subject-area classes mirrors the language arts structure. Adaptation of this structure takes into account the shorter time period of 45 minutes allotted for each subject as well as the unique purposes of each class (e.g., social studies projects, science experiments, practice of math computation). Current research on reading is clearly supportive of teaching adolescents to be strategic (and efferent) in interacting with informational text, primary

sources, and relevant fiction in subject-area classes where comprehension strategies are taught in the service of interpreting text, not as ends in and of themselves (Snow, 2002).

Based on research suggesting that new technology-based models of teaching and learning have the power to dramatically improve educational outcomes (Dede, 1998), the Chicago Striving Readers program is funding various tools to enrich student learning during small-group rotations. Each classroom has:

1. A listening center where students can access models of fluency, record themselves while reading, monitor their own fluency levels, and listen to audio books;
2. A media center with three computers and access to one classroom printer to support special intervention software and group or individual research; and
3. Ten (10) Alpha Smarts (hand-held computers), which are student-friendly alternatives to using paper and pencil for activities such as note-taking, brainstorming, preparing graphic organizers, and taking spelling tests. Alpha Smarts are particularly beneficial to students who have poor handwriting or motor skills or have difficulty in communicating their ideas in writing and organizing their work.<sup>4</sup>

During the 2004–2005 school year, CPS collaborated with university experts to develop three reading instructional approaches that, after proving successful, became integral parts of the Chicago Striving Readers school-wide component:

1. **Direct vocabulary instruction:** This method evolved from CPS’ work with Robert Marzano and the Association for Supervision and Curriculum Development (ASCD). Vocabulary development focuses on a list of subject-specific specialized and technical words, on which direct instruction occurs within regular comprehension instruction and small group activities.
2. **Partner reading for fluency, comprehension, and vocabulary development:** Developed by Dr. Donna Ogle, professor at National-Louis University and senior literacy advisor for the Chicago Striving Readers program, this model provides an opportunity for two students to read aloud to each other and also apply multiple comprehension strategies within a before-during-after framework. This exercise, which can take place during the small-group period in both language arts and subject-area classes, helps students increase reading fluency, improve higher order thinking skills, and build vocabulary.
3. **Text set units related to subject area content and formats:** The Striving Readers text set units in social studies, science, and math increase student motivation and self-directed learning and stimulate the practice of comprehension strategies through guided reading, partner reading, self-selected individual reading, and book club discussions. These units, also developed in collaboration with Dr. Ogle, feature a variety of highly engaging informational non-fiction texts at different reading levels and with varied text structures and organizational features.

---

<sup>4</sup> These computers were distributed to all Striving Readers classrooms by the end of school year 2007-2008, and are currently in use as of the 2008-2009 school year.

## Targeted Schools, Grades, and Students

During its second year of implementation, the Chicago Striving Readers program evaluation involved 63 target schools, including 31 treatment schools (16 from Cohort 1 and 15 from Cohort 2) and 32 control schools (16 per cohort). These schools consist of a variety of grade structures (middle schools, elementary-middle schools and middle-high schools); however, all target schools serve grades 6–8.

Within the treatment schools, all students in grades 6–8 receive at least the whole-school intervention.<sup>5</sup> Students' eligibility for receiving the targeted and/or intensive interventions is based primarily on their reading proficiency in English, as determined by their achievement on the previous school year's standardized reading assessment. Along with other characteristics, reading test results are used to assign students to tiers that determine students' eligibility for services. Students at the highest level of proficiency (Tier 1) are not eligible for targeted or intensive interventions. Students at the next lower level (Tier 2) are assigned to receive the targeted intervention, and those at the lowest level (Tier 3) are assigned to receive the targeted and the intensive interventions. Following is a summary of the procedures for assigning students to tiers that were used during Years 1 and 2 of the program.<sup>6</sup>

In Year 1, students were assigned to tiers based on their Stanford Achievement Test series 10 (SAT-10) scores.<sup>7</sup> For those students with missing SAT-10 test scores, the Basic Reading Inventory (BRI) was administered and the scores from this test were used to generate preliminary tier assignments, which were later modified (as necessary) as soon as SAT-10 data became available. In addition, in some cases, students in treatment schools with SAT-10 scores that were close to the tier cut-off values were also given the BRI as a check on the results of the SAT-10 assessments; those students whose BRI scores indicated a different reading level than the SAT-10 were re-tiered. This policy was not implemented in all schools however.

The tier definitions that CPS program staff had originally established were as follows:

- Students with SAT-10 reading scores at or above grade level would be placed in Tier 1,
- Students with scores below grade level but not more than one year below grade level would be placed in Tier 2, and
- Students who were more than one year below grade level would be placed into Tier 3.

Because these assessments were based on the students' prior year test results, "grade level" for incoming 6th-grade students should have been based on grade equivalents for 5th-grade students. However, for the 2006–2007 school year, tier assignments for these students inadvertently used grade equivalents for students in *6th grade*.<sup>8</sup> As a result, students scoring below the grade equivalent of 5.7 (more than one year below grade level according to the SAT-10 grade equivalents for 6th-grade students) were placed into Tier 3. However, because 5.7 represents

---

<sup>5</sup> With the exception of special education students who are not served in regular classrooms.

<sup>6</sup> Information about these procedures, which took place before Metis Associates' involvement in the evaluation, was obtained from reviews of syntax files from Learning Point Associates (LPA); databases from LPA and CPS; past memoranda from LPA, CPS, and the U.S. Department of Education; and conversations with CPS staff.

<sup>7</sup> Additional criteria were applied to assign ELLs and special education students, which are discussed below.

<sup>8</sup> Memorandum from LPA dated August 22, 2007.

reading *at grade level* for 5th-grade students, this means that in reality all students scoring below grade level were offered intensive intervention services, those reading up to one year *above* grade level were placed into Tier 2, and those reading more than one year above grade level were placed into Tier 1.

CPS program staff identified additional assignment criteria that applied to special education and bilingual/ELL students to ensure that all students in the Striving Readers program could benefit from the services and be capable of meeting project goals. Staff had considered omitting special education students with any disabilities other than learning disability from the program; however, there was no evidence from Year 1 memoranda, nor from tier assignment syntax files,<sup>9</sup> that any special education students were systematically excluded from the tier assignment process (although special education students other than learning disabled *were* removed from the original impact analyses). For English language learners, most of whom did not take the standard statewide literacy exam during the first two program years, the original intention was to place these students in tiers based on their results from the Illinois Measures of Annual Growth in English (IMAGE), an assessment of language development for English language learners. However, IMAGE data were not made available to the district at the time that the Year 1 tier assignments were completed. As a result, only mainstreamed ELL students who were tested on the Illinois Student Achievement Test (ISAT) were systematically assigned to tiers, based on their SAT-10 scores, using the same criteria that were used for English proficient students. However, since some schools received the IMAGE data before it was available to the district, teachers from those schools were able to assign ELL students to tiers based on the intended IMAGE benchmarks, which were assigned as follows: students who tested as *meeting* or *exceeding standards* on the IMAGE were placed in Tier 1, students testing as *below standards* were placed in Tier 2, and those with an IMAGE score categorized as *academic warning* were placed in Tier 3. Other non-mainstreamed ELL students were assigned to tiers once the remaining IMAGE data became available. The BRI was also administered to ELL students, and their tier assignments were adjusted based on these results.

For Year 2 of the program, the same tier definitions were used as in Year 1, but the application of grade equivalents was corrected. The originally-intended tier definitions were thus achieved, with students at or above grade level in Tier 1, those up to one year below grade level in Tier 2, and those more than one year below grade level in Tier 3.<sup>10</sup> The same procedures from Year 1 were also used to assign tiers for students with missing SAT-10 data<sup>11</sup>, special education students, and mainstreamed ELL students. In addition, in Year 2 IMAGE data were available to the district at the time tier assignments were being made, and ELL students who were tested on the IMAGE rather than the ISAT were assigned to tiers based on their IMAGE results.

---

<sup>9</sup> Data and syntax files for Year 1 were obtained from LPA between January and March 2008.

<sup>10</sup> When defining tier assignments for subsequent years, district program developers realized that the intensive intervention is not appropriate for some students who are reading *too* far below grade level. Tier assignments for Year 3 were therefore revised so that only students reading up to three years below grade level were assigned to Tier 3; those with lower reading performance were considered ineligible for Striving Readers services.

<sup>11</sup> In Year 2 of the program, SAT-10 scale scores were not available at the time the tier assignments were made. Instead, tier assignments were based on SAT-10 national percentile ranks (NPR), which were converted into scale scores—and corresponding grade equivalents—using a conversion table from the SAT-10 *Technical Data Report*.

The specific tier assignment criteria used for students entering grade 6 are summarized in Table 1. Assignment criteria for grades 7 and 8 were exactly parallel (based on grade level equivalents) to those for 6th-grade students. In theory, a student could receive targeted and/or intensive intervention services for up to three consecutive years (grades 6 through 8), if they remain eligible according to these criteria. In practice however, this would depend on whether the LIT or another teacher is available to provide these services at the upper grades, which is more often the case in smaller schools.

**Table 1**  
**Tier assignment criteria for entering 6th-grade students**  
**Program Years 1 and 2**

<b>Tier</b>	<b>Assignment Criteria</b>	<b>Year 1<sup>12</sup></b>	<b>Year 2</b>
<b>Tier 1</b>	<b>Grade equivalent</b>	<b>At or above 6.7</b>	<b>At or above 5.7</b>
	SAT-10 NPR* range	--	50–99
	SAT-10 scale score range	≥659	≥643
	IMAGE range (bilingual)	--	≥207
<b>Tier 2</b>	<b>Grade equivalent</b>	<b>5.7 up to 6.7</b>	<b>4.7 up to 5.7</b>
	SAT-10 NPR* range	--	37–49
	SAT-10 scale score range	642–658	629–642
	IMAGE range (bilingual)	--	173–206
<b>Tier 3</b>	<b>Grade equivalent</b>	<b>Less than 5.7</b>	<b>Less than 4.7</b>
	SAT-10 NPR* range	--	1–36
	SAT-10 scale score range	≤641	≤628
	IMAGE range (bilingual)	--	≤172
<b>Students in special education</b>	All special education students with SAT-10 scores: same criteria as general education students.	All special education students with SAT-10 scores: same criteria as general education students.	
<b>Bilingual/ELL students</b>	Mainstreamed ELL students with SAT-10 scores: same criteria as English proficient students.	Mainstreamed ELL students with SAT-10 scores: same criteria as English proficient students.	
	All other ELL students assigned based on BRI results.	ELL students with IMAGE scores: assigned based on cut-off scores specified above.	
<b>Other tier assignment criteria</b>	Students with missing SAT-10 data: preliminary assignment based on BRI results, adjusted once SAT-10 data available.	Students with missing SAT-10 data: preliminary assignment based on BRI results, adjusted once SAT-10 data available.	
	Students who enrolled in the target schools between June and September 2006 were assigned after the second week of school, using the same criteria as above.	Students who enrolled in the target schools after the initial tier assignments were assigned once their test data became available, using the same criteria as above.	

\*NPR = national percentile rank.

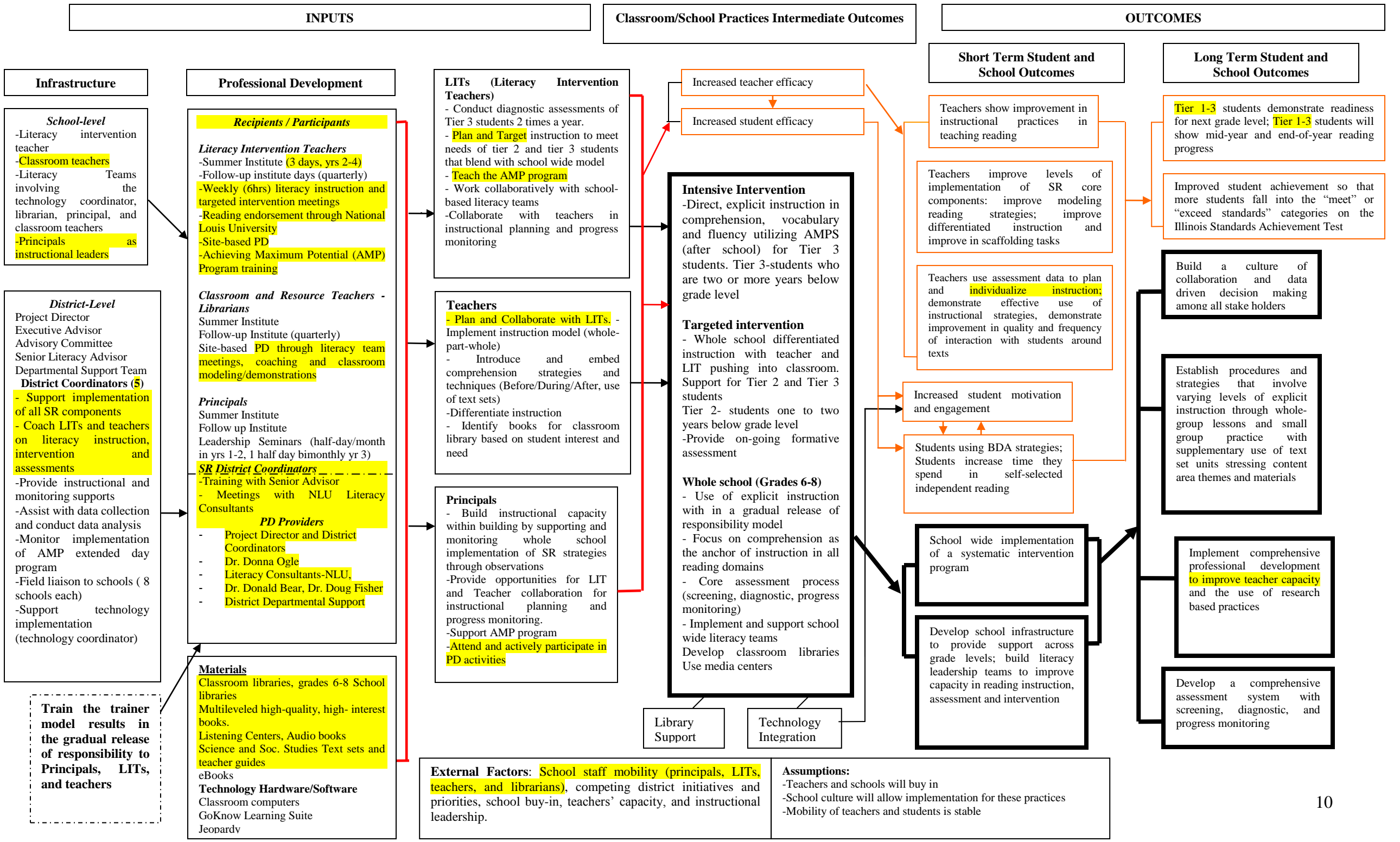
<sup>12</sup> The data for Year 1 represents the *actual* criteria and results. However, it is important to note that these data were not consistent with the *intended* criteria, which would have used grade equivalents one year lower across all tiers.



During the 2007–2008 school year, there were a total of 5,970 students in the intent-to-treat population in grades 6–8 among the 31 Striving Readers schools. Among these, complete data (baseline and posttest reading scores and tier assignments) were available for 4,655 students. Except for the small percentage of students who are served outside of the regular classroom, all of these students received the whole-school blended intervention. In addition, as a result of the tier assignment criteria described above, there were a total of 750 students assigned to Tier 2 who were to receive targeted intervention services, and an additional 1,554 students in Tier 3 who were assigned to receive both targeted and intensive intervention services. The distributions by grade and other demographic characteristics of these students are described further in *Description of First and Second Year Samples* in Section IV.

## **Logic Model**

The logic model for Year 2 of the Chicago Striving Readers program is presented on the following page. This model includes both professional development and classroom model components for both the whole-school and the targeted and intensive interventions. Modifications from the Year 1 model are highlighted.



## Brief Overview of Key Evaluation Design Features

The evaluation of Chicago Striving Readers is a five-year study designed to assess the implementation and impact of the initiative in 31 Title I schools in the Chicago Public School District.<sup>13</sup> Metis Associates, Inc., a national research and consulting firm with offices in New York City, Atlanta, and Philadelphia, was selected by CPS to take over the evaluation in January 2008.<sup>14</sup>

The study examines the implementation and impact of both the *whole-school/blended intervention* model as well as the *targeted* and *intensive intervention* models. The research questions and key design features of the evaluation are summarized below.

### Research Questions<sup>15</sup>

Year 2 of the evaluation used data from a variety of sources to examine the following categories of research questions (the specific research questions are provided in Sections III, IV and V below):

#### 1. Overall Program Impact on Student Achievement

- What is the overall impact of the Chicago Striving Readers program on all participating students' reading scores?
- Is there a differential overall impact of the Chicago Striving Readers program on the reading scores of students at different grades?
- Is there a differential impact of the Chicago Striving Readers program on the reading scores of students in grades 6–8 in different NCLB subgroups, including gender, race, socioeconomic status, and ELL status subgroups?

#### 2. Program Impact on Student Achievement for Struggling Readers<sup>16</sup>

- What is the combined impact of the whole-school, targeted, and intensive interventions of the Chicago Striving Readers program on 6th-grade Tier 3 students' reading scores?<sup>17</sup>
- What is the combined impact of the whole-school and targeted interventions of the Chicago Striving Readers program on 6th-grade Tier 2 students' reading scores?<sup>14</sup>

---

<sup>13</sup> The random assignment process for Cohort 1 schools originally placed 16 schools into the treatment group and 16 into the control group; however, one of the treatment schools did not send a representative to initial program start-up meetings and never became part of the study. The study therefore focuses on 31 treatment schools (15 in Cohort 1 and 16 in Cohort 2).

<sup>14</sup> The process of transitioning from the original evaluator took place in January and February 2008, and the evaluation re-started in March of that year.

<sup>15</sup> These research questions have been modified slightly from Metis's original Evaluation Design Plan of February 2008 to better reflect the current program status.

<sup>16</sup> Because Tier 3 students receive both targeted and intensive interventions, and all students receive the whole-school intervention, it is not possible to isolate the impact of the targeted or intensive interventions, as originally proposed. Instead, program impact is isolated for subgroups of struggling readers.

<sup>17</sup> Because grant resources enable schools to consistently provide targeted and intensive intervention services only at Grade 6, an assessment of the impact of these models at other grades is not supported at present.

- Is there a differential impact of the Chicago Striving Readers intensive intervention program on the reading scores of 6th-grade Tier 3 students in different NCLB subgroups?

### 3. Impact on Classroom Practices

- Were the Chicago Striving Readers schools different than control schools on the seven key program components, including the classroom model (reading comprehension instruction for the whole school, blended intervention; reading comprehension instruction for targeted intervention model for Tier 2 and 3 students; reading comprehension instruction for intensive intervention model for Tier 3 students; purposeful assessment; data-driven instruction; and highly motivating reading materials) and the professional development model?<sup>18</sup>
- Was the Chicago Striving Readers program faithful in its implementation of the proposed program with regard to the seven key components?

An experimental design was established for the assessment of the research questions about program impact. As described further in Section IV, below, this design used random assignment at the school level for each of the two cohorts of schools. Cohort 1 included 32 schools that were randomly assigned to create equal numbers of treatment and control schools that entered the study during the 2006–2007 school year. Cohort 2, which entered the study during school year 2007–2008, originally enrolled an additional 32 schools that were also randomly divided between treatment and control schools. However, one of the Cohort 2 schools that was designated as a treatment school declined to participate in the program, leaving the study with a total of 63 schools. All 63 schools include grades 6–8 among the grades they serve.

The primary measures of student and teacher outcomes that were used during Year 2 of the study include the following:<sup>19</sup>

- **Student Achievement:**
  - Illinois Student Achievement Test (ISAT) – Reading<sup>20</sup>

---

<sup>18</sup> The five original key components, which included professional development; comprehensive assessments; data-driven decision making; high-quality, high-interest materials; and comprehensive instruction, were redefined for the August 2008 Implementation Executive Summary report to create these seven components. One additional component, use of computer-assisted reading instruction, was also added to the evaluation design in the February 2008 proposal and was redefined as Technology Integration (handheld software designed to support small-group differentiated instruction and assessment), which is now a subcomponent of the Targeted Intervention component. However, with approval of the U.S. Department of Education, exploration of this component has been postponed until program Year 3 because implementation of the handheld computers occurred only on a pilot basis during the 2007–2008 school year. Nevertheless, integration of other types of technology-based reading instruction remains part of the focus of the examination of the overall classroom model.

<sup>19</sup> Data collection instruments that were used for the Year 2 study did not obtain information relating to the function and impact of the regional School Coordinators, an important position. Data collection instruments for Year 3 have already been adapted to provide input on the contribution of this role, and additional instruments to obtain input directly from the Coordinators will also be considered.

<sup>20</sup> The original study design also included the Stanford Learning First ClassViews as a second measure of student achievement. However, because it is no longer administered on a regular basis outside of Striving Readers schools, it is no longer possible to use this assessment as a comparison between treatment and control students.

- **Classroom Practices and Program Fidelity:**<sup>21</sup>
  - Adolescent Literacy Observation Protocol
  - Pre-Observation Literacy Environment Checklist
  - School leader interviews (Principals, LITs, Lead Literacy Teachers [LLTs], librarians and technology coordinators)
  - Literacy Improvement Survey for Teachers (LIST)
  - Achieving Maximum Potential (AMP) after-school program schedules
  - AMP attendance records
  - Principal meeting attendance records
  - Teacher professional development schedules
  - Teacher professional development attendance records
  - Teacher professional development agendas

Content and psychometric characteristics of the ISAT, which was used as the primary outcome measure for student achievement, are summarized in the table below (ISBE, 2007; DeStefano *et al.*, 2006).

**Table X**  
**Characteristics of the Illinois Student Achievement Test – Reading**

Grades	Metric	Reading Content Emphasized	Norming Sample and Psychometric Information
3-8	Scale Scores  Performance Levels are derived from the scale scores for each grade level.	ISAT emphases are based on 1985 State Learning Goals and include: <ul style="list-style-type: none"> <li>▪ Read with understanding and fluency.</li> <li>▪ Read and understand literature representative of various societies, eras, and ideas.</li> </ul>	<p><u>Validity:</u> For the purpose of determining Illinois students' relative standing within the national population, the ISAT is equated to the Stanford Achievement Test – Tenth Edition (SAT 10).</p> <p>Correlations with the SAT 10 exceed .94 across the six grades, demonstrating good convergent validity with the nationally accepted norm.</p> <p>Gender- and race-specific confounds were identified using Differential Item Functioning Analysis and systematically replaced until the ISAT functioned comparably for all individuals.</p> <p>Discriminant validity was established using Pearson Separation Values, which exceeded 2.5 in grades 3 through 6 and exceeded 2.1 in grades 7 and 8.</p> <p><u>Reliability:</u> Alpha coefficients for internal consistency, calculated using a parallel testing design, ranged from .86 to .91 for grades 3 through 8.</p> <p>Inter-rater reliability scores for the open-ended response questions exceed 97% for all grade level versions</p>

<sup>21</sup> In addition to the listed measures, the *Surveys of Enacted Curriculum* were also administered to teachers in treatment and control schools. However, this extremely lengthy survey generated very low response rates, particularly from the control schools, and therefore could not be used in fidelity or impact analyses.

### III. Evaluation of the Implementation of the Whole School, Targeted, and Intensive Interventions: Years 1 and 2

---

#### Study Design

##### *Research Questions*

The following research questions relating to assessment of the implementation of the three intervention models (whole-school/blended intervention, targeted intervention, and intensive intervention) were explored during each of the first two years of the Chicago Striving Readers program evaluation.<sup>22</sup>

- **Question 1:** Were the Chicago Striving Readers schools different than control schools on the five [seven<sup>23</sup>] key components: professional development; comprehensive assessments; data-driven decision making; high-quality, high-interest materials; and comprehensive instruction?
- **Question 2:** Was the Chicago Striving Readers program faithful in its implementation of the proposed program with regard to the five [seven<sup>19</sup>] key components: professional development; comprehensive assessments; data-driven decision making; high-quality, high-interest materials; and comprehensive instruction?

##### *Data Sources*

Evaluation of the Chicago Striving Readers initiative uses a mixed-method approach to obtain evidence of program implementation as well as program impact. This approach capitalizes on the different relative advantages of qualitative and quantitative methods so that the findings from parallel measures can be triangulated in order to maximize confidence in the validity of the study's conclusions.

The data collection methods used in the first two years of the study to address research questions relating to assessment of program implementation were summarized in Section II. above. These data sources—many of which served both as evidence of fidelity of program implementation as well as evidence of impact on classroom practices—are described further below.

##### *Observations*

Classroom observations conducted in Years 1 and 2 of the program used a formal observation protocol designed to code all observed activities in 5- to 10-minute intervals. Three waves of observations (fall, winter, and spring) took place during Year 1; in Year 2, as a result of the transition to a new evaluator mid-year, only fall and spring observations were conducted.

---

<sup>22</sup> The February 2008 evaluation proposal for Year 2 had originally added *use of computer-assisted reading instruction* as a sixth key component to be evaluated in the impact study; however, because use of handheld computers occurred only on a pilot basis during the 2007–2008 school year, the impact on this component will be assessed beginning in Year 3 of the program.

<sup>23</sup> See above regarding the change in the definition of the key components.

Observations were conducted in up to three 6th-grade English language arts classrooms in each school. The observations were conducted by trained observers using the structured Adolescent Literacy Observation Protocol (see Appendix A).

### *Surveys*

The Surveys of the Enacted Curriculum (SEC) were administered in the spring of each year to all English language arts teachers who teach students in grades 6, 7, or 8 in treatment and control schools. The English language arts SEC facilitates objective comparisons of what teachers teach to what they are expected to teach. Classroom instruction is compared to state standards, the content of the standards is compared to the content of assessments, and the assessments are compared to instruction. These comparisons measure the teachers' depth of understanding of what is taught, how lessons are articulated across target grades, and how well instruction aligns with state content standards and state assessments.<sup>24</sup>

The Literacy Improvement Survey for Teachers (LIST), including TechPOINT, was also administered each spring to all grade 6–8 English language arts teachers and included sets of items related to various aspects of program implementation.

### *Interviews*

In Years 1 and 2, principals, lead literacy teachers (LLTs), LITs, technology coordinators, and librarians in both treatment and control schools were interviewed twice a year (fall and spring) using a structured interview protocol. District level staff, including the Program Director, senior literacy consultant, and four District Coordinators were also interviewed in winter 2009 regarding the status of program implementation during Year 2.

Table 2 summarizes these data sources, compares those that were used in Year 1 versus Year 2, and shows how each data source was used to address the research questions related to program implementation and impacts on instruction. As this table shows, the Chicago Striving Readers initiative has relied on a broad spectrum of data sources since the beginning of the evaluation. To accommodate correlational analyses of patterns of fidelity of program implementation, as of Year 2, documentation of program activities was maintained at the unit record level wherever possible to facilitate analyses of variations in implementation for different program components and under different circumstances. In addition, because the sheer volume of qualitative and quantitative data sources collected during Year 1 had proven to be impractical to analyze thoroughly, a more targeted data collection plan was established for Year 2 (and beyond) that uses a smaller variety of data sources expected to have the most direct relevance to the study.

---

<sup>24</sup> As previously explained, low response rates prevented the use of this survey in fidelity and impact analyses.

**Table 2**

**Matrix of research questions and data collection methods to assess program implementation: Program Years 1 and 2**

Research Questions: Implementation of Treatment and Impacts on Instruction	Data Sources (# in Target Population for Year 2 Data Collection)																					
	Classroom Observations/ Pre-Observation Checklists (N=60 treatment, 63 control)	Extended-Day Observations (N= up to 63 classes <sup>25</sup> )	Program Leader Interviews (N=6)	Staff (School Leader) Interviews (N=155 treatment, 128 control)	Teacher Lesson Plans	SEC Survey <sup>26</sup> (N=248 treatment, 289 control)	LIST Survey <sup>27</sup> (N=248 treatment, 289 control)	AMP (After-School Program) Schedule (N=31)	AMP Attendance Records (N=31)	Principal Meeting Attendance (N=31)	Literacy Team Agendas	LIT Team Meeting Reflections	Grade Level Meeting Agendas	Professional Development Schedule	Teacher Professional Development Attendance (N=248)	Professional Development Agendas	LIT Time and Effort Reporting Log	School Improvement Plans (SIPAAAs) <sup>28</sup>	Literacy Rich Classrooms	Needs Assessment: Coordinator Observations of Classrooms	Fidelity of Implementation Form Instrument	Summer School Students
<i>Was the Chicago Striving Readers program faithful in its implementation of the proposed program with regard to.../ Were the treatment schools different than control schools on...</i>																						
1. ...professional development?	Y1			X	X		X	X			X		X		X	X	X		X	X		X
	Y2			X	X		X	X			X		..		X	X	X		..	..		..
2. ...comprehensive assessments?	Y1	X	X	X	X	X	X	X			X	X	X				X	X			X	
	Y2	X	X	X	X	..	X	X			X	..	..				..	..			..	
3. ...data-driven decision making?	Y1			X	X	X	X	X			X	X	X				X	X			X	
	Y2			X	X	..	X	X			X	..	..				..	..			..	
4. ...high-quality, high-interest materials?	Y1	X	X	X	X	X	X	X			X	X	X				X	X		X	X	
	Y2	X	X	X	X	..	X	X			X	..	..				..	..		..	..	
5. ...comprehensive instruction?	Y1	X	X	X	X	X	X	X	X		X	X	X				X	X		X	X	X
	Y2	X	X	X	X	..	X	X	X	X		X	..	..			..	..		..	..	..

<sup>25</sup> Most schools' AMP/after school programs had ended before the spring observations took place.

<sup>26</sup> The SEC was administered in the spring of Year 2 but the response rate was too low for the results to be useable.

<sup>27</sup> The LIST was administered in the spring of Year 2. Response rates from control schools were too low to use the results for assessment of differences in program impact on instruction between treatment and control schools, but the results were used to assess program implementation in the treatment schools.

<sup>28</sup> These documents will be collected again beginning with the Year 3 evaluation.



## Year 1 Implementation Study

### *Intervention as Implemented*<sup>29</sup>

Although the program model is differentiated by design to meet the varying needs of different schools, classrooms, staff, and students, successful implementation of the Striving Readers Initiative requires that participating schools implement certain critical components of the model at least at a minimum level. The role that these key components play in moving towards programmatic goals and objectives was represented graphically in the logic model presented in Section II, which also provided a summary of the key program components.

The following tables summarize the findings, where available, regarding variations in fidelity of implementation of each of the first three key program components of the classroom model during Year 1, as originally described in the Year 1 Implementation Report. It should be noted that, as was often also the case with evidence about implementation in Year 2 (as reported below), the nature of the data collection instruments made it difficult to assign evidence about implementation to specific intervention models, because the phrasing of these instruments often was not explicit about the context. For this reason, most of this evidence should be interpreted as reflecting fidelity of implementation of the overall Chicago Striving Readers program, rather than any particular intervention model. (Data that did provide explicit evidence relevant to particular intervention models are described in Table 4 [targeted intervention] and Table 5 [intensive intervention].) For Year 3, this problem is being corrected by modifying the data collection instruments so that they address whether particular activities, strategies, and resources are taking place during small-group instruction, during the AMP program, or during regular classroom activities.

---

<sup>29</sup> As part of the August 2008 Executive Summary of the Year 1 Implementation Report, fidelity scales were developed by Metis Associates in collaboration with CPS, in order to provide a quantifiable summary of the fidelity of program implementation. However, because of the transition of the study to a new evaluator as of January 2008, these fidelity scales were not applied to data collected prior to that transition. Summaries of variations in the fidelity of program implementation for Year 1 that are reported here were obtained from Learning Point Associates (*Chicago Striving Readers Implementation Report: Year 1 (2006–2007)*, 2007, Chicago, IL).

**Table 3**  
**Variability in implementation of Striving Readers instructional activities and strategies:**  
**Overall program implementation**

Program Features	Summary of Status of Intervention, Year 1
<p><b>PRC2; text sets and technology integration are used fluidly and alternately to support differentiated instruction and increase student motivation, engagement, and understanding.</b></p> <p><u>Partner Reading in the Content Area, Too (PRC2):</u> a reading instructional framework to support reading comprehension and fluency of nonfiction text</p> <p><u>Text sets:</u> high interest books used to help students read strategically, promote engagement and motivation and deepen their content knowledge.</p> <p><u>Technology integration:</u> integration of classroom computers, listening centers and other technology materials designed to support small group differentiated instruction.</p>	<p>LITs were trained on PRC2 and reported being prepared to immediately introduce partner reading when possible. However, text materials necessary for PRC2 were unavailable in Year 1. As such, content-related literacy support through this avenue had not yet begun.</p> <p>Text set were made available to students participating in the whole-school/blended intervention through integration into the regular curriculum and through the school libraries. There was some disconnect, however, between availability and use of these materials: according to summaries of classroom observations in Year 1, these materials were only being used approximately one third of the time (i.e., such materials were used during 33% of the “potentially relevant opportunities”).</p> <p>Technological resources were already implemented fairly regularly during Year 1. Listening centers were used at least once a week by half (51%) of the teachers, and less than once a week but more than once a month by an additional 28% of teachers. Media centers were used by 60% of the teachers in the fall of 2006 and 85% of the teachers in the spring of 2007.</p>
<p><b>Whole-part-whole instructional framework</b></p>	<p>Roughly half of all teachers participating in the Striving Readers program (from 44% to 53%) were observed using a whole-part-whole instructional framework. On average, however, these teachers only spent a “moderate” amount of time doing so. Furthermore, findings were ambiguous as to whether entire literacy blocks were structured around this framework and the extent to which whole-group and small-group instructional activities were integrated and connected.</p>
<p><b>Use of gradual release model to provide direct, explicit instruction and scaffold learning for students</b></p>	<p>Insufficient data were reported from Year 1 to explicitly determine whether a comprehensive gradual release model was being used. Nevertheless, important components of the model were reported to have been observed fairly regularly. For example, coaching and scaffolding in small-group activities took place in more than two thirds (69%) of small-group activities observed in spring 2007.</p>

**Table 3 (cont.)**  
**Variability in implementation of Striving Readers instructional activities and strategies:**  
**Overall program implementation**

Program Features	Summary of Status of Intervention, Year 1
<p><b>Instruction anchor for all classrooms and content areas is focused on comprehension</b></p>	<p>Although a majority of teachers reported that they regularly covered at least some of the comprehension strategies, it is unclear whether coverage was consistently sustained for all seven core comprehensive strategies. What follows is the percentage of teachers reporting that they regularly covered each strategy<sup>30</sup>: The majority of teachers reported sustained coverage of <i>prediction</i> and <i>metacognitive strategies</i> (74% and 63%, respectively). However, only about one third of respondents reported “sustained” or “considerable” coverage of <i>questioning strategies</i> and <i>inferring</i> (38% and 32%, respectively), and only 14% reported “considerable” coverage of <i>summarization</i> strategies. No information was available from the Year 1 reports regarding the degree to which teachers implemented comprehension strategies relating to <i>visualization</i> or <i>text structure</i>.</p>
<p><b>Highly motivating reading materials integrated with engaging technology and audio resources</b></p>	<p>Although there were some delays in integrating some of the program-related technology, tremendous strides were made during Year 1. The vast majority (87%) of interviewed staff reported that technology was at least <i>somewhat integrated</i> by spring 2007. However, only two fifths (40%) of those interviewed felt that technology was <i>thoroughly integrated</i>. The majority of teachers (65%) also reported spending at least <i>some</i> instructional time using computers and technology to learn, practice, or explore language arts, although only two fifths (39%) reported spending at least <i>moderate</i> instructional time using computers and technology.</p> <p>However, delays in acquisition and distribution of some materials created limitations. According to the Year 1 report, “classroom observations suggest... that even once materials were in place, they were not being fully used. Looking across three of the major types of materials targeted for use by the program—text sets, listening centers, and computer media centers—classroom observations show a use of these resources in about one third of the classrooms.”</p>
<p><b>Frequent assessment and adjustment of instruction</b></p>	<p>The large majority (86%) of teachers in the Striving Readers schools reported that their instructional practices have been positively influenced by diagnostic and assessment results, and two thirds (66%) also indicated that they worked with their schools’ LLTs to “use assessment data for instructional planning.” A solid majority (74%) of Striving Readers teachers also reported that they “use assessments to directly inform and drive instruction,” and 76% reported that district-level tests had at least a <i>somewhat positive</i> influence on what they teach (although only 29% reported that they had a <i>strong positive influence</i>). Assessment data—including from Learning First, ISAT, BRI, and informal assessment—were used for a variety of instructional purposes, from screening and benchmarking to assessing outcomes. The majority (66%) of teachers, however, did not use Running Records as part of their instruction. Some additional setbacks to the use of assessments also occurred as a result of an initial delay in the hiring of the LITs from March to June 2006, which prevented them from receiving training (including orientation to the use of the BRI) prior to the start of the program and prolonged the initial period of assessment.</p>

<sup>30</sup> Coverage of the comprehension strategies was rated on two separate scales. Some were rated on a scale of *none/little/some/moderate/considerable*, while others were rated on a scale of *none/slight/moderate/sustained*.

**Table 3 (cont.)**  
**Variability in implementation of Striving Readers instructional activities and strategies:**  
**Overall program implementation**

<p><b>Direct/explicit vocabulary instruction:</b>  Systematic approach to teaching academic content vocabulary in all subjects using Robert Marzano’s <i>Building Academic Content Vocabulary</i></p>	<p>Approximately one fifth to one third of all observed English language arts classes (27%, 31%, and 19% of observed classes during fall, winter, and spring observations, respectively) included small-group activities focusing on vocabulary. Two fifths (40%) of surveyed teachers felt that the Striving Readers program was <i>very effective</i> in helping them develop the use of partner reading for vocabulary development, and almost another two fifths (38%) felt it was at least <i>moderately effective</i>. Evidence from Year 1 was not sufficient, however, to determine whether vocabulary instruction was systematic or content focused, or whether it specifically used Marzano’s techniques.</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Table 4**  
**Variability in implementation of Striving Readers instructional activities and strategies:**  
**Targeted intervention**

Program Features	Summary of Status of Intervention, Year 1
<p><b>Teachers and LITs collaborate in instructional planning and progress monitoring</b></p>	<p>Collaborative working relationships between classroom teachers and LITs were fostered through initial professional development activities. These staff specifically collaborated on the incorporation of assessment-informed instructional planning; by winter 2007, all LITs reported actively collaborating with teachers—primarily “helping teachers use their assessment data to group their students and to inform and drive instruction.” Through the collaboration process, the role of the LITs transformed from that of advisor to that of a peer. For example, in the fall of 2006, the primary role of the LIT was to administer the assessments while in the spring of 2007, the primary role of the LIT was to “collaborate with teachers on curriculum.” Teachers and LITs also had the opportunity to collaborate through two types of instructional teams. Literacy Leadership Teams, which work at the school-level to ensure the program is on track, were meeting weekly in 63% of the schools by spring 2007, and grade-level teams were meeting weekly in 67% of the schools by this time.</p>
<p><b>Increased direct and supported instruction:</b>  Approximately 20–30 minutes per day within a 60- to 90-minute language arts block</p>	<p>Direct teaching of targeted youth during 20-minute periods of small-group instruction by LITs took place in 46% to 59% of the literacy blocks that were observed during Year 1. However, it is likely that other classrooms were also using supported instruction at times other than when the observations took place.</p>
<p><b>Explicit instruction in seven core comprehension strategies</b></p>	<p>See Table 3, above, regarding overall focus on comprehension strategies. No evidence was available specific to the implementation of these strategies during the small-group activities.</p>

**Table 5**  
**Variability in implementation of Striving Readers instructional activities and strategies:**  
**Intensive intervention**

Program Features	Summary of Status of Intervention, Year 1
<b>Increased time:</b> An additional 240 minutes of direct and supported instruction beyond the intervention that occurs during the regular school day	Additional direct and supported literacy instruction was provided through the AMP after-school program for Tier 3 students. Increased time was therefore systematically provided when these programs were offered and when students attended. AMP classes generally met from October 2006 through April or early May 2007. Student attendance in the AMP classes throughout Year 1, averaging 81%, was reasonably high but might not have been high enough to ensure that all Tier 3 students received the full weekly average of 240 minutes of additional instruction.
<b>Small-group setting:</b> 15:1 ratio of students to teacher	The intensive intervention took place in settings that exceeded the goal of an average 15:1 student-teacher ratio. Across all classes, there was an average of one teacher for every 10 students (although some individual schools ran programs with more than 15 students per teacher).
<b>Explicit and systematic instruction in seven core comprehension strategies:</b> Strategies introduced one at a time	See Table 3, above, regarding overall focus on comprehension strategies. No evidence was available specific to the implementation of these strategies in the AMP classes.
<b>Teaching of high volume and depth of academic vocabulary</b>	See Table 3, above, regarding overall focus on vocabulary instruction. No evidence was available specific to the implementation of this focus in the AMP classes.
<b>Guided fluency practice</b>	Guided fluency practice was provided to Tier 3 students through the AMP program. AMP includes vocabulary building, fluency, word identification skills, and background knowledge through interactive and diagnostic-based computer software, and the AMP program was implemented in every after-school program. The value of the AMP reading materials may have been hindered by the extent to which students were engaged with them: LITs noted that some of the topics were of limited relevance and that others became tedious for students when spread over several days.

### *Implications for Impact Analyses*

It is to be expected that any large, complex instructional program involving systemic change at the school and district level would take some time to get up to speed. A relevant example is found in a report by Horizon Research, Inc., which has been studying local systemic change initiatives in math and science for many years. Findings in that report addressing the likelihood of institutionalization over time indicate that it takes at least two years (Banilower, Boyd, Pasley, & Weiss, 2006). The report also showed the over-time trajectory of three “systems” necessary for district-wide adoption: “systems for professional development,” “systems for aligning district policies,” and “systems for garnering and maintaining stakeholder support.” According to this study, these systems also do not reach equilibrium before the completion of the second year. The Chicago Striving Readers Initiative is no exception.

As indicated from the Year 1 findings summarized above, there were a number of program components that were not yet fully in place—and in some cases, not in place at all—during the first program year, while others had a slow start early in the year. This could result from delays in starting the training of LITs or in obtaining materials and resources, as well as from the time that teachers require to become comfortable using particular instructional strategies and methods.

Indeed, literature on the adoption of systemic change initiatives has shown that it can take several years for the change to become fully integrated.

Nevertheless, Chicago's Striving Readers initiative got off to a strong start on a number of important dimensions. As summarized above, the AMP schedule was fully implemented during Year 1 and, overall, teachers were already reporting wide use of assessments for instructional planning. Most teachers also reported using computers to support instruction, although in many cases technology was not yet thoroughly integrated. There was also widely reported use of instruction in at least some of the seven comprehension strategies, although this was considerably more limited for *questioning* and *summarization*. Although teachers reported feeling ready to implement the PRC2 instructional method, the necessary materials were not acquired during the first program year. Perhaps the most notable obstacle was the delay in the hiring of Literacy Intervention Teachers, who play such a central role in both the targeted and intensive instruction models, but who began their training even as they were beginning to fulfill their responsibilities. In addition (and possibly related to the challenge of LITs beginning their jobs while they were being oriented), there was limited observation of small-group instruction by LITs and only "moderate" time using the whole-part-whole method according to self-reports. Finally, although the program started on time, student attendance at AMP classes could have been better, and it was not clear whether enrollment was aligned with tier assignments.

Obviously, to the extent that any program components did not get started or were not fully integrated from early in the year, analyses would be expected to show limited (if any) impact on student achievement in the first year. Similarly, for the current analyses, although the Cohort 1 students should have had at least as much exposure to the intervention as those in Cohort 2 during Year 2, the additional dosage of exposure that would normally be expected from their having started a year earlier would be attenuated by these delays. As a result, it would be less likely to expect that cohort would serve as a significant explanatory variable in the Year 2 analyses of overall program impact.<sup>31</sup>

## Year 2 Implementation Study

### *Intervention as Implemented*

Based on the data sources used to provide evidence about key program features, as discussed above, a series of rubrics were created that were used to generate scores representing level of program implementation.<sup>32</sup> In addition, interviews with district-level program leadership were used to further illuminate some of the quantitative results of the fidelity rubrics. Rubric scores represent the adequacy with which the program has been implemented for a particular school, classroom, or demographic group relative to the original program model. They are generated by comparing actual versus intended levels of implementation on factors such as the following:

---

<sup>31</sup> Because analyses of impacts specific to Tiers 2 and 3 included only 6th-grade students, this phenomenon would not have affected the degree to which cohort might serve as an explanatory variable in these analyses, since Year 1 variations in dosage for grade 6 students would only affect outcomes for grade 7 students in Year 2.

<sup>32</sup> Because of the transition to a new evaluator during the second program year, these scores could not be calculated for Year 1.

- **Professional development model** – proportion of targeted staff attending trainings and amount of different types of training received.
- **Classroom model** – degree of emphasis on key instructional and assessment components, proportion of targeted students receiving targeted instruction, amount of time students receive intensive instruction (AMP after-school program attendance), availability of resources, extent of librarian support, extent of integration of technology and other subject areas into literacy instruction, etc.

It should be noted that, depending on the source of implementation data, it was possible to calculate some scales at the grade, tier, or classroom level, but others could only be measured at the school level. However, because total scores for each major program component were derived by aggregating across data sources, final scores were only calculated at the school level. All scores except the professional development scale were defined on a 10-point scale, where a 1 indicates that none of the key program characteristics are being implemented (according to observations and self-reports on surveys and interviews), and a 10 indicates that all key components were being implemented. (The professional development scale starts at 0, which represents a low average rate of attendance at all of the key training activities.) It should be noted, however, that a 10 does not represent a “perfect” score. It is always theoretically possible for any school, classroom, or teacher to do more; however, scale values were capped at 10, which represents implementation as defined by the model. Further details about the fidelity scale definitions are presented in Appendix B.<sup>33</sup>

CPS interprets fidelity scores ranging from 8 to 10 as representing *high implementation*, scores above 5 but less than 8 as representing *medium implementation*, and scores of 5 or lower as representing *low implementation*. Results of the fidelity scales for Year 2 program implementation are presented below.

---

<sup>33</sup> As explained above, SEC data were not incorporated into these scales due to low response rates.

**Table 6**  
**Results of Year 2 Implementation Fidelity Scales**

<b>Fidelity Component</b>	<b>Cohort</b>	<b>N</b>	<b>Mean Score</b>	<b>Standard Deviation</b>	<b>Median Score</b>	<b>Minimum Score</b>	<b>Maximum Score</b>
<b>Overall fidelity</b>	Cohort 1	16	6.6	0.36	6.6	6.0	7.2
	Cohort 2	15	6.6	0.41	6.6	5.7	7.4
	<b>Total</b>	<b>31</b>	<b>6.6</b>	<b>0.38</b>	<b>6.6</b>	<b>5.7</b>	<b>7.4</b>
Component 1: Whole-school (blended) intervention	Cohort 1	16	7.6	0.44	7.6	6.9	8.8
	Cohort 2	15	7.1	0.59	7.0	5.8	8.5
	<b>Total</b>	<b>31</b>	<b>7.3</b>	<b>0.58</b>	<b>7.4</b>	<b>5.8</b>	<b>8.8</b>
<i>Sub-Component 1: Whole-part-whole</i>	Cohort 1	16	9.1	0.75	9.2	7.5	10.0
	Cohort 2	15	8.6	1.05	8.8	6.3	10.0
	<b>Total</b>	<b>31</b>	<b>8.9</b>	<b>0.93</b>	<b>9.1</b>	<b>6.3</b>	<b>10.0</b>
<i>Sub-Component 2: Gradual release model</i>	Cohort 1	16	8.6	0.49	8.7	7.7	9.5
	Cohort 2	15	8.3	0.47	8.3	7.5	9.1
	<b>Total</b>	<b>31</b>	<b>8.5</b>	<b>0.49</b>	<b>8.5</b>	<b>7.5</b>	<b>9.5</b>
<i>Sub-Component 3: Comprehension focus</i>	Cohort 1	16	5.3	0.61	5.2	4.2	6.2
	Cohort 2	15	5.1	0.69	5.1	4.1	6.2
	<b>Total</b>	<b>31</b>	<b>5.2</b>	<b>0.64</b>	<b>5.2</b>	<b>4.1</b>	<b>6.2</b>
<i>Sub-Component 4: PRC2</i>	Cohort 1	16	7.3	0.85	7.4	5.9	8.6
	Cohort 2	15	6.8	0.91	6.4	5.6	8.2
	<b>Total</b>	<b>31</b>	<b>7.1</b>	<b>0.91</b>	<b>7.1</b>	<b>5.6</b>	<b>8.6</b>
<i>Sub-Component 5: Marzano's vocabulary</i>	Cohort 1	16	7.6	1.16	7.6	4.9	9.4
	Cohort 2	15	6.5	1.65	6.7	3.8	9.7
	<b>Total</b>	<b>31</b>	<b>7.0</b>	<b>1.50</b>	<b>7.3</b>	<b>3.8</b>	<b>9.7</b>
Component 2: Targeted Intervention	Cohort 1	16	5.9	1.05	5.9	3.9	7.6
	Cohort 2	15	5.9	1.02	5.9	4.2	7.8
	<b>Total</b>	<b>31</b>	<b>5.9</b>	<b>1.01</b>	<b>5.9</b>	<b>3.9</b>	<b>7.8</b>
<i>Sub-Component 6: Teacher/LIT collaboration</i>	Cohort 1	16	6.5	2.09	6.0	2.0	10.0
	Cohort 2	15	6.7	1.79	6.0	4.0	10.0
	<b>Total</b>	<b>31</b>	<b>6.6</b>	<b>1.92</b>	<b>6.0</b>	<b>2.0</b>	<b>10.0</b>
<i>Sub-Component 7: Direct instruction in comprehension</i>	Cohort 1	16	5.3	0.61	5.2	4.2	6.2
	Cohort 2	15	5.1	0.69	5.1	4.1	6.2
	<b>Total</b>	<b>31</b>	<b>5.2</b>	<b>0.64</b>	<b>5.2</b>	<b>4.1</b>	<b>6.2</b>
Component 3: Intensive Intervention	Cohort 1	16	8.1	0.85	8.1	6.6	9.9
	Cohort 2	15	7.5	1.30	7.5	4.9	10.0
	<b>Total</b>	<b>31</b>	<b>7.8</b>	<b>1.13</b>	<b>7.9</b>	<b>4.9</b>	<b>10.0</b>
<i>Sub-Component 8: Increased instructional time</i>	Cohort 1	16	6.8	1.56	6.8	3.2	9.7
	Cohort 2	15	5.1	2.32	5.0	1.9	10.0
	<b>Total</b>	<b>31</b>	<b>6.0</b>	<b>2.12</b>	<b>6.0</b>	<b>1.9</b>	<b>10.0</b>
<i>Sub-Component 9: Small-group setting</i>	Cohort 1	16	9.5	0.88	10.0	7.3	10.0
	Cohort 2	15	9.8	0.56	10.0	7.9	10.0
	<b>Total</b>	<b>31</b>	<b>9.6</b>	<b>0.75</b>	<b>10.0</b>	<b>7.3</b>	<b>10.0</b>
Components 4 & 5: Purposeful Assessment & Data Driven Instruction	Cohort 1	16	6.9	0.78	7.0	5.7	8.6
	Cohort 2	15	6.8	0.84	6.7	5.0	8.3
	<b>Total</b>	<b>31</b>	<b>6.8</b>	<b>0.80</b>	<b>6.8</b>	<b>5.0</b>	<b>8.6</b>
Component 6: Materials	Cohort 1	16	6.5	0.69	6.6	5.3	7.8
	Cohort 2	15	6.0	0.62	5.9	5.0	7.2
	<b>Total</b>	<b>31</b>	<b>6.3</b>	<b>0.70</b>	<b>6.2</b>	<b>5.0</b>	<b>7.8</b>
Component 7: Professional development	Cohort 1	16	4.8	1.19	4.7	2.7	6.7
	Cohort 2	15	6.2	1.30	6.7	4.0	8.0
	<b>Total</b>	<b>31</b>	<b>5.5</b>	<b>1.43</b>	<b>5.3</b>	<b>2.7</b>	<b>8.0</b>



As these data show, all schools in both cohorts were implementing the Striving Readers program at a medium level of overall fidelity during the 2007–2008 program year, with no overall difference between cohorts. While variations in scores for different scales would seem to indicate that certain program components were implemented more successfully than others, such comparisons must be interpreted with caution, since the extent to which the scales fully reflect all aspects of the program model varies from component to component. For example, at the component level, the highest mean ratings were observed for the intensive intervention (Component 3). However, the fidelity scales as defined only reflect the extent to which additional instructional time was provided to Tier 3 students and the extent to which instruction was provided in small-group settings. In addition, data sources relating to other aspects of program implementation were not explicit about which model they applied to. For example, survey questions about the use of instructional methods (such as direct vocabulary instruction) that are part of the intensive intervention model did not ask respondents to distinguish between their implementation of such activities during the regular class period or during the AMP program. Thus, the scores for Component 3 do not reflect either the completeness or the quality of implementation of the *content* of the intensive intervention classes. Indeed, as was also discussed in reference to the Year 1 implementation, both LITs and district staff have expressed concerns about the content of the AMP program, which has been described as having limited relevance for some students and limited adaptability.

Similar limitations restrict the extent to which Component 2 reflects the content of the targeted intervention. For these reasons, these scales might best be interpreted as reflecting *overall* program fidelity.<sup>34</sup> Fidelity of specific program components should be interpreted within the limitations of the focus of the data sources on which they are based.

With these caveats about the limitations of the fidelity scales as currently defined, the most successful measured program component was the creation of small-group settings during the after-school program—all schools in both cohorts met, or at least came very close, to the required 15:1 student-teacher ratio. Nevertheless, there was still considerable variability by school, several of which did not reach the “high implementing” level of fidelity on this sub-component.

Among the specific instructional methods that make up the Chicago Striving Readers program, the highest mean fidelity ratings were obtained for use of the whole-part-whole and the gradual release model, both of which had mean ratings across all schools at the “high implementing” level. Again, it should be noted that a major component of the original fidelity scale for whole-part-whole was to be based on classroom observation codes; since there was no clear way to interpret whether patterns of activities coded in 5-minute intervals really represented a whole-

---

<sup>34</sup> The relatively limited range of fidelity scores among schools on most components also does not reflect district staff’s impressions of the actual range of program implementation. In order to determine the extent to which such discrepancies may be due in part to the design of the fidelity scales and/or the implementation rubrics used by program staff, fidelity scores will be compared to results of the implementation rubric and to staff’s impressions of implementation by school and by component. The implications of any gaps in alignment will be studied in order to determine whether any modifications are needed in the data collection instruments to improve their alignment with the program model and to increase the validity of the fidelity scales for Year 3.

part-whole structure, this component was instead scored based only on LIST survey responses. As a result, both of the highest ratings were for scales for which ratings were based only on self-report. Although there is no reason to expect that responding teachers would consciously distort their ratings, their accuracy also depends on the respondents' level of understanding of the methods that they are rating (a concern which has been echoed by several members of the project's district leadership team). It is also notable that, despite the high overall score for the whole-part-whole model, several schools were only rated in the "medium implementing" level of fidelity on this scale, with the lowest ratings, among Cohort 2 schools, at the low end of this range.

The next highest fidelity ratings were found for implementation of the gradual release model. In this case again, none of the original data sources used during Year 2 provided explicit evidence for implementation of this model; however, the LIST survey did ask questions about the use of instructional practices that are important components of the model, including scaffolding, differentiated instruction, guided reading, and monitoring comprehension through questioning. Almost all schools ranked in the "high implementing" level of fidelity on this sub-component, while only one Cohort 1 school and three Cohort 2 schools fell slightly short of that mark.

The program components where implementation appeared to be the most problematic included anchoring instruction in comprehension, explicit instruction in comprehension during small-group activities and increased instructional time for the after-school program. For all three of these program sub-components, mean scores across schools were at the low end of a "medium implementing" level of fidelity. For both of the comprehension sub-components, five of 16 Cohort 1 schools and seven of 15 Cohort 2 schools fell into the "low implementing" range; while for increased instructional time, two of 16 Cohort 1 schools and eight of 15 Cohort 2 schools fell into the "low implementing" range.

Implementation varied more dramatically from school to school on some program components than others. Most notable were ratings of collaboration between the LIT and classroom teachers and increased instructional time for the intensive intervention, both of which ranged from perfect or near perfect scores to the low end of the "low implementing" level of fidelity. In addition, rather substantial variations among schools were observed for direct vocabulary instruction and implementation of whole-part-whole instructional structures.

The professional development component of the fidelity scales (Component 7) was also rated at a "medium implementing" level of fidelity for most schools; however, it was on a lower end of this range than most of the other components relevant to the classroom model, and for a number of schools—including nine of the 16 Cohort 1 schools and three of the 15 Cohort 2 schools—it fell into the "low implementing" range. This fidelity scale did not reflect all aspects of the initiative's professional development program, but it was designed to rate the schools on their level of participation in the core elements of training. Low scores on this scale mean that principals, LITs, and/or teachers had low average attendance rates at at least some of the following training sessions:

- LIT weekly meetings with coordinators
- Principals' monthly professional development

- Teachers' summer institute
- Teachers' monthly Saturday seminars
- Teachers' quarterly follow-up institutes

The Project Director noted that it is particularly difficult for many teachers to attend the Saturday seminars, and this may be a significant reason for the lower fidelity scores on this component. The Director suggested that additional on-site assistance for classroom teachers could provide an important source of support to help make up for this gap. A substantial number of the respondents to the spring 2008 school leader interviews cited the importance of the professional development program to the Striving Readers Initiative, particularly for helping them to integrate literacy into other content areas, and staff from the Striving Readers schools concurred that professional development was one of the main strengths of the literacy curriculum. Respondents also expressed a desire for more training and more support, although a number of respondents cited a lack of buy-in among teachers as a challenge to implementation. Attendance problems at professional development activities are also further exacerbated by turnover among teachers, which—although formal data could not be obtained in time for this report—district staff have noted is high in many Striving Readers schools. Although the Coordinators have observed that new teachers are usually excited about the initiative, they have expressed concerns about the disadvantage that new teachers have due to having missed a portion of the training. Although the District Coordinators and the LITs already strive to target their support where it is needed most, their efforts can be stretched thin, especially in larger schools. An increase in the availability of on-site technical assistance, if possible, could also help provide a source of additional support for new teachers who have not received the entire compliment of training. A few principals also commented that some of the professional development activities that they were expected to attend went into more detail than necessary—but on the whole the comments about the Striving Readers professional development program were very positive, and Striving Readers schools reported more participation in training around literacy than control schools.

Because the interview data were not disaggregated by cohort, it is difficult to find an explanation in these data for why Cohort 1 schools had lower attendance rates. Certainly it was the intent of the initiative that all of these activities should continue into the second program year. It is possible that the lower attendance among Cohort 1 personnel reflects a feeling that they had already received sufficient training in some areas—or that they found the training less useful than did Cohort 2 participants. On the other hand, it might simply be that Cohort 1 schools are encountering greater scheduling problems. Given the importance of the training program to the initiative, it may be worthwhile to survey staff more directly about their perceptions of professional development and the factors influencing their attendance.

Results of the fidelity scales are presented by school in Appendix C.

In addition to measures of implementation fidelity obtained through surveys of school-based staff, interviews of district-level program staff were conducted during winter 2009 in order to obtain a perspective from project leadership on the Initiative's implementation status during its second year. Among those interviewed were the Program Director, the Senior Literacy Advisor, and the four District Coordinators, results of which are discussed throughout this report.

Additional insights into certain key program components—beyond the numerical results of the fidelity scales—emerged from these interviews. Although the program components of purposeful assessment and data driven instruction received only medium fidelity ratings on average, a closer look provides a clearer image of how the Striving Readers community is learning to use such data. The fidelity scores for this component reflect numerous facets of the efforts to use assessment data, including whether the use of data for instructional planning occurs through collaboration with a school literacy coach, during grade-level team meetings and/or during literacy team meetings. It also reflects to what purposes different assessment data are used and to what extent. It is possible therefore that schools can be fairly strong in a number of these areas, but still receive modest fidelity scores if they are missing some of these facets.

Project leadership affirm that the program was making substantial progress in major aspects of this objective during Year 2. They speak of creating a “community of learners” that, through the venue of literacy team meetings and with the critical support of the LITs, use student assessment data to assign students to targeted intervention models, individualize and differentiate instruction, group students for small group activities (such as during Whole-Part-Whole and PRC2 activities), determine class level progress, and to purchase and select appropriate reading materials. This habit of mind, according to the Coordinators, has also led to an increase in data-driven program management, using data from LIT reflections, which provide them “a good understanding of where more support is needed.”

Project leaders also spoke to the Initiative’s use of technology to support instruction. Although the handheld computers had not yet been widely distributed until the end of Year 2 and technology use was not included as a formal objective until Year 3, the use of classroom media centers and listening centers formed a significant part of the program design from its inception. District staff explained that the Striving Readers schools had a mixture of successes and challenges in their use of technology during Year 2. Many teachers see benefits in the use of technology to support instruction, although even among those who do, it often poses a significant challenge in requiring them to learn the applications and, perhaps more significantly, to adapt their instructional methods.

The classroom media centers, for example, are most often used for small group work, such as for conducting research on a project during group work conducted as part of whole-part-whole activities. A question was raised, however, as to whether these resources were being used for independent work—which could limit, in part, the extent to which they can be used to differentiate instruction. Use of the listening centers was limited somewhat during Year 2 due to delays in acquiring some of the audio books and associated software. While these delays may have temporarily limited the usefulness of the listening centers in helping students access models of fluency, one District Coordinator pointed out that they could still be used to allow students to record themselves and listen to themselves read. That teachers were not always accessing this feature of the centers was seen as an indication that teachers still needed more professional development in the use of these technologies.

Additional insights into the fidelity and variability of program implementation were derived from interviews with principals, LITs, LLTs, librarians, and technology coordinators. A detailed

summary of these findings, including school leaders' perceptions of the professional development program, are presented in Appendix D.

### ***Implications for Impact Analyses***

Because several of the data sources used to measure implementation provided incomplete measures of fidelity, it is difficult to infer the implications of these results for program impacts. This is especially true regarding expected impacts of the targeted and intensive interventions, since most fidelity measures did not explicitly focus on program characteristics specific to these models. Nevertheless, certain generalizations can be made. The finding that overall program implementation, on average, was rated at the “medium level” of program fidelity is perhaps consistent with where one would expect a comprehensive five-year reform initiative to be in its second year. For this very reason, however, it is unlikely that the full potential of the program to impact student achievement would be observable at this stage. To the extent that certain essential program components, such as the focus on comprehension strategies, may have been slow in getting started, it would be expected to reduce the chances of achieving impact even further. Even where impediments to quality or dosage of implementation occurred only at particular schools—such as with instructional time in AMP and extent of small-group instruction—it would still reduce the likelihood of being able to measure program impact, since these impacts are measured across all schools.

### **Comparison of Year 1 and Year 2 Implementation**

Because of the transition between evaluators and changes in the definitions of key program components from Year 1 to Year 2, it was not practical to generate fidelity scales for Year 1 data. As a result, direct comparisons of implementation were not possible. Nevertheless, certain generalizations could be inferred by comparing implementation summaries from the Year 1 report (summarized in Tables 3 through 5 above) to results of the Year 2 fidelity scales (summarized in Table 6). Among program components that were reported to have gotten off to a slow start in Year 1, the use of PRC2 showed the most dramatic improvement. Although teachers reported feeling adequately prepared to implement this technique, the necessary materials had not been acquired and this component was not implemented in Year 1. By Year 2, however, materials had been distributed and this program component received one of the highest overall fidelity ratings. Nevertheless, as noted in the discussion of fidelity scales above, they may have been imperfect measures of implementation quality since they rely heavily on self-report; in this case, the district team believes that PRC2 was still at the beginning level of implementation during Year 2 because, while implemented regularly, it was often not being implemented correctly.

The whole-part-whole instructional model was reported to have occurred in about half of observed 6th-grade language arts classes during Year 1. By Year 2, whole-part-whole instruction was rated the highest among all program sub-components. Comparisons are complicated by the fact that the Year 1 report did not specify what pattern of observation codes were interpreted as representing whole-part-whole instruction; the results may have simply focused on the use of small-group instruction, as was the case in Year 2. Nevertheless, the fidelity code for Year 2 also incorporated teachers' self-reports in response to the question, “How often do you use... [the] whole-part-whole classroom instruction model?” Thus, the high

fidelity score means that most teachers reported relatively frequent use of this specific instructional model.

In the first program year, teachers' responses to the SEC revealed some limitations in the extent of use of at least some of the core comprehension strategies. Although low SEC response rates in Year 2 prevented direct comparisons, it is notable that self-reports (from the LIST survey) about the extent to which instructional methods designed to help struggling readers develop comprehension resulted in fidelity ratings toward the low end of "medium implementation." However, because these scales reflect instructional methods promoted by Striving Readers that are designed to improve comprehension<sup>35</sup> as well as specific comprehension strategies,<sup>36</sup> it is not clear whether the lower ratings reflect a lack of a strong focus on comprehension or just lack of use of specific methods.

Finally, one area that proved to be a challenge to program implementation in both years was increased instructional time for Tier 3 students. The amount of additional time was limited somewhat in Year 1 by irregular attendance for some students. A similar pattern occurred in Year 2, where irregular attendance (combined in some cases with AMP classes that only met for one hour after school) resulted in Cohort 2 schools receiving the lowest rating for this sub-component of any of the fidelity scales.

Interviews with district staff revealed a consensus among project leaders that continuity in reform efforts is critical to their success. Unfortunately, for a variety of reasons, such continuity was maintained with only limited success during the second program year. For example, the Senior Literacy Advisor has been a constant and essential source of support since the Chicago Striving Readers Initiative was first conceptualized; and those schools that were able to continue working with the same District Coordinator since the first year have found value in this continuity. In contrast however, a number of unexpected challenges that arose during the second year put a damper on the initiative's progress and necessitated the continued focus on transitional efforts. Among the major factors leading to this condition were the replacement of several leadership positions (most notably the Project Director, as well as the Technology Director and three of the District Coordinators), as well as some delays in acquisition of some technology and literacy resources. These transitional conditions were further exacerbated by a sense that competition exists for the Coordinators' and Project Director's time, and for resources among the district's numerous instructional initiatives. While substantial progress was made during the second year of the initiative, these challenges slowed the rate of that progress, and the program was in many ways still in the developmental stage during its second year. As a result, there remained a wide range of perceptions of the program among school personnel, from those who already grasped the initiative as an effort to achieve systemic change with a focus on improving comprehension, to others who still saw the program from a more concrete perspective as an assortment of innovative instructional methods.

---

<sup>35</sup> Such as *Exclusion Brainstorming* and *Guided Reading and Summarizing Procedure (GRASP)*

<sup>36</sup> Such as understanding the arrangement of text (*text structure*), making connections to background knowledge and making connections between texts (*metacognition*), and synthesizing information within text or across texts (*summarizing*).



## IV. Evaluation of the Impact of the Targeted and Intensive Interventions: Years 1 and 2

---

### Study Design

#### *Research Questions*

The research questions relating to assessment of the impacts of the targeted and intensive intervention models that were explored during the first two years of the evaluation of the Chicago Striving Readers initiative are presented in Table 7, below. In Year 2, research questions relating to exploring the initiative’s differential impact on students in different demographic subgroups were added to the study.<sup>37</sup>

**Table 7**  
**Research questions relating to impact of the targeted and intensive interventions**

Year 1	Year 2
1) What is the impact of the Chicago Striving Readers program on 6th-grade Tier 3 students’ reading scores (i.e., intensive intervention impact), as measured by ClassViews and the ISAT?	1) What is the impact of the Chicago Striving Readers program <sup>38</sup> on 6th-grade Tier 3 students’ reading scores, as measured by the ISAT <sup>39</sup> ?
--	2) Is there a differential impact of the Chicago Striving Readers program <sup>32</sup> on the reading scores of 6th-grade Tier 3 students in different NCLB subgroups, including gender, race, socioeconomic status (as determined by eligibility for free/reduced-price lunch), and ELL status, as measured by ISAT <sup>33</sup> ?

To more fully represent the program design, Metis staff conferred with the federal grant consultant, Abt Associates, and agreed that analyses in addition to the research questions in Table 7 should be conducted to explore program impact on the reading scores of 6th-grade Tier 2 students.

#### *Sampling Plan*

To facilitate an unbiased assessment of the impacts of Striving Readers, the study randomly assigned participating schools to either a treatment or control condition. As described in the Year 1 Impact Report<sup>40</sup>:

*The process of random assignment began with CPS generating a list of schools having a high percentage of struggling readers, as measured by their performance on a*

<sup>37</sup> Research questions relating to the program’s impact on classroom practices are discussed in Section V.

<sup>38</sup> The original wording of these research questions described them as exploring the “intensive intervention impact”; however, because Tier 3 students are also supported by the whole-school and targeted intervention models, the evaluation of the program impact on this particular group of students cannot isolate the impact of any particular intervention model.

<sup>39</sup> As discussed above, ClassViews was not used as an outcome measure in Year 2 because it is no longer consistently administered.

<sup>40</sup> Learning Point Associates, December 2007.



*standardized test. These schools were invited to participate in the program and informed that by agreeing to participate, they also were agreeing to be either a treatment or a control school. Control schools were promised a stipend of \$5,000 per year for their assistance with the evaluation. From this pool of eligible and willing participants, treatment and control schools were randomly selected by Learning Point Associates. Willingness to participate in the activities of the program and the evaluation was confirmed in writing during an informational session run by the CPS Office of Literacy.*

As further detailed in Appendix A of the Year 1 report, a power analysis established that a minimum sample of 32 schools (16 treatment and 16 control) would achieve adequate power (approximately 85%) to detect an effect size of roughly one-third standard deviation change over time with  $\alpha = 0.05$ . Random assignment of schools occurred over two years, creating two cohorts of schools. Thirty-two schools were randomly assigned to either the treatment or control condition in school year 2006–2007 (Cohort 1), and another 32 schools were randomly assigned in 2007–2008 (Cohort 2). As previously discussed however, one of the Cohort 2 schools assigned to the treatment group did not join the study, so the resulting sample includes a total of 31 treatment (Striving Readers) and 32 control schools.

The intent-to-treat model used for these impact analyses includes all students in grades 6–8 in the 31 treatment schools who were eligible for services according to the Year 2 tier assignment criteria (described in greater detail in Section II, above) and who were enrolled in the target schools at any time during the 2006-2007 school year (for students in Cohort 1 schools) or during the 2007-2008 school year (for students in Cohort 2 schools)<sup>41</sup>. Among the control schools, all students in these same grades who have matched (pre and post) ISAT scores, including tested ELL and special education students, are included in the comparison group. (Although English language learners in treatment schools who were tested only on the IMAGE test were also assigned to tiers in Year 2 based on the results of this test, these students could not be included in the impact analyses because they do not have the same outcome measure.<sup>42</sup>)

In order to further control for pre-program differences between students in treatment and control schools, outcome analyses were conducted by tier, comparing results for treatment students to control students who met the same criteria that were used to assign the treatment students to tiers, as described in Section II. In Striving Readers schools, students in Tier 2 receive the whole-school intervention with additional support through scaffolded instruction that is provided during the regular classroom reading block (the targeted intervention). The students in Tier 3, the lowest performing students, receive the intensive intervention (the AMP after-school program) in addition to the targeted and whole-school interventions.

---

<sup>41</sup> With the exception of students who crossed over between treatment and control schools. These students would be assigned to the group in which they originated for the purpose of the ITT analyses. For Cohort 2, this was determined by comparing the spring 2008 ISAT database file to the fall 2007 official enrollment file (September 20th file). For Cohort 1, only the spring 2007 ISAT file was used to identify students because the fall 2006 enrollment file had not been provided to the evaluator.

<sup>42</sup> Since Chicago began administering both the ISAT and the district's new ELL test, the ACCESS, to ELL students as of spring 2008, these students will be included in impact analyses beginning in Year 3.

### Sample Size and Power

There were a total of 4,655 Striving Readers students and 4,688 control students in the Year 2 study with complete data: 2,642 and 2,655 students, respectively, in Cohort 1, and 2,013 and 2,033 students, respectively, in Cohort 2. Among these students, those assigned to Tiers 2 or 3 receive the targeted intervention, and those assigned to Tier 3 also receive the intensive intervention. The distribution by tier is presented in Table 8.

**Table 8**  
**Sample size for Year 2 analyses by cohort, treatment group, and tier<sup>43</sup>**

Cohort Group	Treatment Group	Total N	Tier 1		Tier 2		Tier 3	
			N	%	N	%	N	%
Cohort 1	Control	2,655	1,444	54%	412	16%	799	30%
	Treatment	2,642	1,549	59%	372	14%	721	27%
	Total	5,297	2,993	57%	784	15%	1,520	29%
Cohort 2	Control	2,033	753	37%	312	15%	968	48%
	Treatment	2,013	802	40%	378	19%	833	41%
	Total	4,046	1,555	38%	690	17%	1,801	45%
All students	Control	4,688	2,197	47%	724	15%	1,767	38%
	Treatment	4,655	2,351	51%	750	16%	1,554	33%
	Total	9,343	4,548	49%	1,474	16%	3,321	36%

In order to determine the probability of detecting treatment effects in the tested population, power analyses were conducted for each hierarchical linear modeling (HLM) model that was conducted. Power analyses were conducted using *Optimal Design* software for cluster randomized trials.<sup>44</sup> All power analyses used the following assumptions:

- Two-level HLM model (student and school)
- Type I error rate (alpha) = 0.05
- Intra-class correlation (rho) = 0.10
- Number of clusters (schools) = 63

In each case, the power to detect an effect size of at least .33, and the smallest effect size that could be detected at an ideal power level of 80%, were determined. Results of the power analyses are presented below.

#### **Model 1: Grade 6 Tier 2 (Whole-School Plus Targeted Interventions)**

- Additional assumption: Cluster size (n) = 7 (average number of grade 6 Tier 2 students per school)

#### **Results:**

Effect size = .333      Power = 79%  
Effect size = .341      Power = 80% (ideal power level)

<sup>43</sup> Distribution of intent-to-treat students with complete data.

<sup>44</sup> Retrieved from [http://sitemaker.umich.edu/group-based/optimal\\_design\\_software](http://sitemaker.umich.edu/group-based/optimal_design_software)

### ***Model 2: Grade 6 Tier 3 (Whole School Plus Targeted Plus Intensive Interventions)***

- Additional assumption: Cluster size (n) = 18 (average number of grade 6 Tier 3 students per school)

#### Results:

Effect size = .333      Power = 92%

Effect size = .274      Power = 80% (ideal power level)

As these results show, the distribution of students and schools in this study provides sufficient statistical power to detect a program effect of as little as one third standard deviation, while the model designed to assess the impact on grade 6, Tier 3 students has sufficient power to detect program effects of about one quarter (.27) standard deviation.

#### ***Data Collection Plan***

The main outcome measure for student impact is reading scores on the ISAT, a statewide exam administered to all students in Chicago in grades 6–8 in the spring of each year.<sup>45</sup> The 2008 ISAT in reading is comprised of items from the SAT-10, published by Pearson Education, and items written by Illinois teachers. Pearson Education has created a standard, norm-referenced abbreviated version of the SAT-10 assessment (SAT 10/Abb), which comprises 30 items that are also consistent with Illinois State Frameworks. The inclusion of these 30 SAT-10 items on the ISAT exam allows for the reporting of nationally norm-referenced results such as national percentile ranks (ISBE, 2008). The overall ISAT exam in reading is made up of multiple-choice items (51 items in each grade for grades 5 through 8) as well as one extended-response item for each grade. Scoring of the ISAT is based on the number of correct items, with weighting based on item difficulty. Because test items change each year, the Rasch model of item response theory is applied to the ISAT multiple-choice tests in order to equate the scores from year to year and thereby create a vertically scaled test that allows for longitudinal comparisons. Details of the test development can be found in the technical manual (ISBE, 2007).

Students' ISAT achievement in reading is measured using performance levels of *exceeds standards*, *meets standards*, *below standards*, and *academic warning*. All ISAT reading scores are also reported on a continuous vertical scale from grade 3 to grade 8. Although the minimum score has been set at 120 for all grades and subjects, the maximum score is determined based on each year's ISAT data and may change from year to year. The following table (which is taken from Table 2 of ISBE's *Illinois Standards Achievement Exam: Guide to the 2008 Illinois State Assessment*, presents the scale score ranges for reading performance for grades 3–8.

---

<sup>45</sup> Prior to school year 2007–2008, English language learners in Illinois took a different language test (the IMAGE) in place of the ISAT. As of spring 2008, however, all ELL students took the ISAT in addition to the ACCESS.

**Table 9**  
**Scale score ranges defining student performance levels on the 2008 ISAT scales**

Grade	Academic Warning	Below Standards	Meets Standards	Exceeds Standards
3	120–155	156–190	191–226	227+
4	120–157	158–202	203–236	237+
5	120–160	161–214	215–246	247+
6	120–166	167–219	220–256	257+
7	120–173	174–225	226–266	267+
8	120–179	180–230	231–277	278+

### *Summary of Analytic Approach to the Impact Analysis*

Hierarchical linear modeling was used to assess the impact of the targeted and intensive interventions on students’ reading performance. The impact analyses focused on the intent-to-treat populations (see description of samples later in this section). Because school was the unit of assignment but impacts are measured at the student level, two-level models were used for these cross-sectional analyses in order to account for the clustering of students in schools.

This section presents the results of three different sets of analyses. Because there is no group of students in the Chicago Striving Readers schools who receive either the targeted or intensive intervention exclusively (all students at least also receive the whole-school intervention, and Tier 3 students receive all three intervention models), it is not possible to isolate the impacts of any one intervention model. Rather, the analytic approach was designed to isolate the impacts of the Chicago Striving Readers initiative as implemented on specific groups of students. Thus, the first set of analyses was designed to assess the combined impact of the whole-school and targeted interventions on Tier 2 students. For analyses of impacts on Tier 2 and 3 students focused on the 6th grade. LITs’ availability was less consistent in the higher grade classrooms in some schools—particularly in grade 8. Specifically, the program focus expanded into grade 7 in Year 2, but while all Cohort 2 classrooms were supported by the LIT, 7th grade classrooms in three of the 16 Cohort 1 schools were not. Because program services had not yet expanded into grade 8 in Year 2, 8th grade classrooms were supported by the LIT in only 4 of 16 Cohort 1 schools and 8 of 15 Cohort 2 schools. As a result, the targeted intervention was not implemented as consistently in grades 7 and 8, and for this reason, only sixth-grade students were included in these analyses.

The second set of analyses was designed to address research question 1, which relates to the program’s impact on 6th-grade Tier 3 students. In addition, a third set of analyses was designed to assess research question 2, which relates to whether Striving Readers had a differential impact on specific subgroups of students within Tier 3. Results from this set of analyses are presented under *Additional Analyses*, below.

As shown in Table E-1 in Appendix E, a number of covariates were included in the full models to assess the program impact on 6th-grade Tier 2 and Tier 3 students. At the student level, the models included the following covariates: Black, Hispanic, baseline reading score, baseline math score, gender, special education status<sup>46</sup>, and eligibility for free or reduced-price lunch. English

---

<sup>46</sup> Identified as students with an Individualized Education Plan (IEP).

proficiency was not included as a covariate in Level 1 because the samples used in these analyses had very few English language learners with valid baseline ISAT scores.<sup>47</sup> At the school level, the full models included the following variables: proportion of minority (non-White) students, proportion of students reading at or above grade level, proportion of students at or above grade level in mathematics, proportion of special education students, proportion of ELLs, proportion of students eligible for free/reduced-price lunch, school size in targeted grades, cohort, and treatment. The analyses examining the differential impact of Striving Readers on specific subgroups of students also included interaction terms between treatment and these subgroups.

Appendix E describes in greater detail the approaches used to fit these models, the specification of the models, the selection of covariates, and the treatment of missing data. An overall description of the sample used for these analyses is provided below.

## **Description of First- and Second-Year Samples**

### ***Numbers and Basic Characteristics of Schools, Teachers, Classrooms, and Students***

Because the proportion of students in a school who possess certain characteristics can influence the school environment in ways that affect all students over and above the effects that these characteristics might have on the individual students who possess them, we examined school size and the proportion of students in each school (within the grades targeted by Striving Readers) who were female, English language learners, special education, or low income. Among other reasons, total enrollment can be a significant factor for a Striving Readers program because it determines the number of students for whom each school's LIT is responsible. School characteristics by cohort and treatment group are presented in the following five tables.

---

<sup>47</sup> Prior to the 2006–2007 school year, ELLs in Illinois were expected to take the IMAGE test instead of the ISAT to assess their English proficiency (in 2006–2007, the IMAGE test was replaced by the ACCESS test). In 2007–2008, ELLs were also required by the Illinois State Board of Education to take the ISAT test.

**Table 10a**  
**School characteristics: Cohort 1 control schools**

School Number	Target Grades 6-8 for School Year 2007-2008											
	Total N	% Female	% LEP	% Special Education	% Low Income	% Attendance	% Race					
							Amer. Indian	Asian	African American	Hispanic	White	Other/Multiracial
2	218	55%	7%	0%	86%	95.1%	0%	0%	97%	0%	0%	2%
3	173	39%	13%	0%	98%	91.8%	0%	0%	100%	0%	0%	0%
5	114	50%	23%	2%	83%	94.3%	1%	0%	76%	21%	0%	2%
12	152	41%	14%	0%	97%	91.7%	0%	0%	99%	1%	0%	1%
14	300	53%	12%	1%	82%	93.6%	0%	0%	98%	1%	0%	1%
15	94	45%	14%	24%	98%	96.6%	0%	0%	1%	97%	1%	1%
16	361	48%	10%	14%	96%	92.8%	0%	0%	59%	37%	0%	3%
19	162	44%	11%	15%	94%	95.0%	1%	0%	13%	77%	4%	5%
21	411	48%	11%	14%	88%	95.1%	0%	0%	1%	86%	10%	3%
23	380	50%	12%	17%	79%	95.8%	0%	2%	1%	68%	24%	3%
26	51	59%	12%	0%	94%	91.8%	0%	0%	90%	2%	4%	4%
27	228	54%	12%	29%	99%	96.3%	0%	0%	1%	99%	0%	0%
28	223	46%	9%	22%	84%	95.9%	0%	8%	40%	28%	19%	5%
29	141	62%	14%	0%	78%	95.8%	0%	0%	96%	0%	1%	2%
30	89	58%	11%	0%	88%	94.5%	0%	0%	99%	0%	1%	0%
31	168	43%	15%	14%	96%	94.8%	0%	0%	5%	85%	2%	8%

**Table 10b**  
**School characteristics: Cohort 1 treatment schools**

School Number	Target Grades 6-8 for School Year 2007-2008											
	Total N	% Female	% LEP	% Special Education	% Low Income	% Attendance	% Race					
							Amer. Indian	Asian	African American	Hispanic	White	Other/Multiracial
4	102	51%	15%	4%	86%	95.2%	2%	0%	38%	51%	2%	7%
5	176	51%	10%	10%	98%	95.4%	0%	1%	2%	89%	6%	3%
6	323	49%	12%	16%	100%	97.9%	0%	0%	4%	94%	1%	1%
8	230	52%	17%	0%	89%	94.7%	0%	0%	97%	0%	0%	3%
11	585	54%	9%	12%	97%	95.9%	0%	0%	5%	91%	3%	1%
13	108	50%	13%	0%	94%	93.0%	0%	0%	100%	0%	0%	0%
16	475	53%	14%	16%	85%	95.3%	0%	5%	2%	64%	22%	7%
17	103	54%	17%	0%	99%	95.3%	0%	0%	100%	0%	0%	0%
19	219	53%	18%	13%	93%	95.6%	0%	1%	0%	89%	4%	6%
20	179	43%	18%	1%	82%	95.0%	0%	0%	88%	7%	2%	3%
22	227	46%	10%	15%	84%	95.2%	0%	0%	7%	85%	4%	4%
24	69	43%	29%	1%	97%	94.1%	0%	0%	97%	3%	0%	0%
27	101	50%	8%	16%	87%	95.9%	0%	0%	12%	87%	0%	1%
29	134	47%	16%	9%	90%	95.4%	1%	0%	9%	76%	5%	9%
30	98	51%	27%	52%	98%	95.6%	0%	0%	0%	99%	1%	0%
31	249	43%	12%	22%	91%	96.2%	0%	18%	6%	59%	12%	5%

**Table 10c**  
**School characteristics: Cohort 2 control schools**

School Number	Target Grades 6-8 for School Year 2007-2008											
	Total N	% Female	% LEP	% Special Education	% Low Income	% Attendance	% Race					
							Amer. Indian	Asian	African American	Hispanic	White	Other/Multiracial
1	75	48%	11%	0%	100%	93.6%	0%	0%	100%	0%	0%	0%
4	185	50%	22%	9%	92%	92.3%	0%	0%	45%	50%	1%	4%
6	117	53%	15%	1%	97%	90.4%	0%	1%	97%	1%	0%	1%
7	101	50%	15%	4%	97%	93.0%	1%	0%	89%	8%	0%	2%
8	207	52%	16%	0%	98%	94.5%	0%	0%	99%	0%	0%	0%
9	176	51%	9%	0%	98%	93.0%	0%	0%	100%	0%	0%	0%
10	162	41%	17%	0%	96%	92.2%	0%	0%	98%	0%	1%	2%
11	143	52%	17%	0%	97%	91.8%	0%	0%	100%	0%	0%	0%
13	165	55%	12%	1%	99%	91.2%	0%	0%	99%	0%	0%	1%
17	96	54%	26%	0%	95%	90.8%	0%	0%	100%	0%	0%	0%
18	243	53%	11%	0%	97%	93.4%	0%	0%	99%	0%	0%	1%
20	76	47%	14%	8%	100%	89.6%	0%	0%	80%	20%	0%	0%
22	122	49%	12%	7%	94%	94.5%	0%	0%	80%	17%	2%	2%
24	81	53%	19%	0%	98%	89.2%	0%	0%	99%	0%	1%	0%
25	138	37%	20%	0%	92%	92.8%	0%	0%	99%	0%	0%	1%
32	166	47%	20%	1%	95%	92.4%	0%	0%	99%	1%	0%	1%



**Table 10d**  
**School characteristics: Cohort 2 treatment schools**

School Number	Target Grades 6-8 for School Year 2007-2008											
	Total N	% Female	% LEP	% Special Education	% Low Income	% Attendance	% Race					
							Amer. Indian	Asian	African American	Hispanic	White	Other/Multiracial
1	42	48%	29%	0%	98%	93.1%	0%	0%	90%	5%	2%	2%
2	147	52%	11%	0%	95%	93.3%	0%	0%	99%	0%	0%	1%
3	128	46%	13%	0%	100%	90.3%	0%	0%	100%	0%	0%	0%
7	94	53%	18%	1%	73%	95.7%	0%	0%	97%	2%	0%	1%
9	232	53%	15%	0%	95%	93.3%	0%	0%	99%	0%	0%	0%
10	110	46%	23%	0%	99%	91.2%	0%	0%	95%	2%	0%	3%
12	129	48%	16%	1%	98%	92.2%	0%	0%	97%	2%	1%	1%
14	145	48%	14%	6%	95%	93.0%	0%	3%	68%	23%	2%	4%
15	238	51%	16%	0%	95%	94.3%	0%	0%	99%	0%	0%	0%
18	130	53%	18%	0%	85%	90.2%	0%	0%	99%	0%	0%	1%
21	120	47%	22%	1%	99%	92.5%	0%	0%	100%	0%	0%	0%
23	79	41%	10%	0%	100%	91.3%	0%	0%	100%	0%	0%	0%
25	157	48%	23%	1%	92%	93.4%	0%	0%	99%	1%	0%	0%
26	124	43%	19%	0%	98%	95.9%	1%	0%	98%	0%	2%	0%
28	154	49%	12%	0%	97%	91.5%	0%	0%	99%	0%	0%	1%

**Table 10e**  
**School characteristics:**  
**Summary by Treatment Group by Cohort**

Group	Cohort	Target Grades 6-8 for School Year 2007-2008											
		Total N	% Female	% LEP	% Special Education	% Low Income	% Attendance	% Race					
								Amer. Indian	Asian	African American	Hispanic	White	Other/Multiracial
Control	1	3,265	49%	12%	11%	89%	94.4%	0%	1%	47%	44%	6%	3%
	2	2,253	50%	16%	2%	96%	92.2%	0%	0%	92%	6%	0%	1%
Treatment	1	3,378	50%	14%	12%	92%	95.4%	0%	2%	24%	65%	6%	3%
	2	2,029	49%	16%	1%	95%	92.7%	0%	0%	96%	2%	0%	1%

The demographic characteristics of Striving Readers and control students who were described in Table 8 are summarized in Table 11 for those students in Tiers 2 and 3 (struggling readers). This table provides demographic characteristics of all students, across tiers, and across both cohorts. As these data show, within their populations of striving readers, the treatment and control groups were demographically very similar, although some notable differences were apparent. One difference between Cohort 1 and Cohort 2 was that the latter schools were selected from an applicant pool that represented African American students much more heavily and Hispanic students less heavily than the Cohort 1 schools.<sup>48</sup> In addition, students' racial distribution leaned more heavily toward African American students among Cohort 1 control schools, while Cohort 1 treatment schools were more heavily Hispanic. Otherwise, the treatment and control schools from in both cohorts remained very similar to each other.

---

<sup>48</sup> This was also associated with a much smaller proportion of ELL students in Cohort 2 schools; however, because the process of assigning students to tiers excludes most ELL students, this pattern isn't apparent in this table.

**Table 11**  
**Characteristics of intent-to-treat Tier 2 and Tier 3 students with complete data\***

Characteristics		All Students		Cohort 1 Students		Cohort 2 Students	
		Control (N=2,491)	Treatment (N=2,304)	Control (N=1,211)	Treatment (N=1,093)	Control (N=1,280)	Treatment (N=1,211)
Grade level	6	32%	34%	30%	30%	34%	37%
	7	35%	34%	36%	38%	33%	31%
	8	33%	32%	34%	32%	32%	32%
Gender	Female	45%	44%	45%	43%	46%	46%
	Male	55%	56%	55%	57%	54%	54%
Race/Ethnicity	American Indian	0%	0%	0%	0%	0%	0%
	Asian	0%	0%	0%	1%	0%	0%
	African American	79%	70%	62%	40%	95%	97%
	Hispanic	18%	25%	33%	51%	4%	1%
	White	1%	2%	2%	4%	0%	1%
	Other/Multiracial	2%	2%	2%	3%	1%	1%
% Special education <sup>49</sup>		23%	26%	22%	28%	24%	24%
<b>Disabilities:</b>							
	504	0.0%	0.1%	0.0%	0.2%	0.0%	0.0%
	Autistic	0.2%	0.3%	0.2%	0.4%	0.3%	0.2%
	Educational and Behavioral Disorder	1.5%	1.7%	1.3%	2.1%	1.6%	1.3%
	Educable Mentally Handicapped	2.0%	2.2%	1.2%	1.4%	2.8%	3.0%
	Learning Disorder	14.5%	17.4%	15.8%	19.6%	13.3%	15.4%
	Mentally Handicapped	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%
	Trainable Mentally Handicapped	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
	Other Health Impaired	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%
	Physically Handicapped	0.1%	0.0%	0.2%	0.0%	0.0%	0.0%
	Speech/Language	0.7%	0.6%	0.7%	0.6%	0.8%	0.6%
	Traumatic Brain Injury	0.0%	0.1%	0.1%	0.1%	0.0%	0.2%
	Visually Impaired	0.1%	0.0%	0.2%	0.0%	0.0%	0.0%
	Hearing Impaired	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%
	Multiple diagnoses	0.1%	0.2%	0.2%	0.4%	0.0%	0.0%
	Unknown	3.6%	3.1%	2.4%	0.0%	4.8%	3.3%
% English language learners		2%	2%	3%	3%	1%	0%
% Free/reduced-price lunch eligible		95%	95%	92%	95%	98%	96%

\*Baseline scores, tier assignments, and outcome data were available for all students in this analysis.

## Impacts on Students at the End of Two Years

Table 12 shows the impact estimates of Striving Readers on 6th-grade Tier 2 students (who received the targeted intervention) and on 6th-grade Tier 3 students (who received both the targeted and intensive interventions).<sup>50</sup>

<sup>49</sup> Special education students were eligible for program services and would have been assigned to tiers using the BRI if they did not have an ISAT score; however, only those with ISAT scores could be included in impact analyses. Reasons for students being omitted from analyses, including missing test scores, are broken down in Appendix E.

<sup>50</sup> Impact estimates are based on simple models that did not include interaction terms. The full and final models that were used to calculate these estimates are presented in Tables F-2 and F-3 (for Tier 2 students) and Table F-4 and F-

**Table 12**  
**Estimated impact of Striving Readers on Tier 2 and Tier 3 students**

Population Group	Unadjusted Means		Regression-Adjusted Means		Estimated Impact	Effect Size	p Value
	Control	Treatment	Control	Treatment			
Grade 6 Tier 2 students	224.33	225.85	223.92	226.61	2.69	.20	.085
Grade 6 Tier 3 students	206.35	209.01	207.48	209.21	1.74	.10	.168

As shown in Table 12, for both analyses, students in the treatment group had slightly higher mean reading scale scores than students in the control group; however, the differences were not statistically significant and the effect sizes were small (.20 and .10, respectively). These results indicate that there was no detectable *overall* impact of the program on Tier 2 and 3 students among the intent-to-treat sample. However, further examination of the covariates in the model provides a better understanding of the differential program impacts for various demographic subgroups. These results are discussed below.

### **Additional Analyses**

As noted above, the evaluation was also designed to assess whether there was a differential impact of the Striving Readers program on the reading performance of different NCLB subgroups among 6th-grade Tier 3 students. To assess this, variables representing the different subgroups were entered into the models and their interactions with treatment were explored. Results are presented in Tables 13 and 14, below, which present results for the full model and for the final model (after removing covariates with *p* values greater than .20), respectively.

---

5 (for Tier 3 students) in Appendix F.

**Table 13**  
**Full model for program impact on grade 6 Tier 3, with interactions**

<b>Fixed Effects</b>	<b>Coefficient</b>	<b>SE</b>	<b>p</b>	<b>Effect Size</b>
<b>Model for mean school reading achievement (B<sub>0</sub>)</b>				
Intercept (G <sub>00</sub> )	207.509	0.942	.000	--
PMIN (G <sub>01</sub> )	6.611	18.462	.721	0.383
PREAD07 (G <sub>02</sub> )	17.578	7.525	.023	1.018
PSPED (G <sub>03</sub> )	7.609	16.056	.637	0.440
PLEP (G <sub>04</sub> )	-4.991	13.110	.705	-0.289
PLUNCH (G <sub>05</sub> )	-20.341	13.486	.137	-1.177
SIZE08 (G <sub>06</sub> )	0.002	0.008	.762	0.000
COHORT (G <sub>07</sub> )	-1.988	1.762	.265	-0.115
TRT (G <sub>08</sub> )	1.731	1.329	.199	0.100
<b>Model for BLACK Slope (B<sub>1</sub>)</b>				
Intercept (G <sub>10</sub> )	-0.154	4.161	.971	-0.009
TRT (G <sub>11</sub> )	0.211	5.203	.968	0.012
<b>Model for HISPANIC Slope (B<sub>2</sub>)</b>				
Intercept (G <sub>20</sub> )	1.852	4.281	.665	0.107
TRT (G <sub>21</sub> )	-3.308	5.490	.547	-0.191
<b>Model for BASEISAT Slope (B<sub>3</sub>)</b>				
Intercept (G <sub>30</sub> )	0.428	0.044	.000	0.025
TRT (G <sub>31</sub> )	0.050	0.065	.447	0.003
<b>Model for BASEMATH Slope (B<sub>4</sub>)</b>				
Intercept (G <sub>40</sub> )	0.159	0.045	.001	0.009
TRT (G <sub>41</sub> )	0.011	0.064	.860	0.001
<b>Model for GENDER Slope (B<sub>5</sub>)</b>				
Intercept (G <sub>50</sub> )	-1.616	1.198	.178	-0.094
TRT (G <sub>51</sub> )	-0.969	1.730	.575	-0.056
<b>Model for IEP Slope (B<sub>6</sub>)</b>				
Intercept (G <sub>60</sub> )	-9.746	1.517	.000	-0.564
TRT (G <sub>61</sub> )	-0.151	2.185	.945	-0.009
<b>Model for LUNCH Slope (B<sub>7</sub>)</b>				
Intercept (G <sub>70</sub> )	3.028	3.478	.384	0.175
TRT (G <sub>71</sub> )	4.078	4.514	.367	0.236
<b>Random Effects (Var. Components)</b>	<b>Variance</b>	<b>df</b>	<b>Chi-Square (p)</b>	<b>ICC</b>
Var. in school means (U <sub>0</sub> )	11.804	54	114.554 (.000)	0.059
Var. within schools (R)	188.482			

**Table 14**  
**Final model for program impact on grade 6 Tier 3, with interactions**

<b>Fixed Effects</b>	<b>Coefficient</b>	<b>SE</b>	<b>p</b>	<b>Effect Size</b>
<b>Model for mean school reading achievement (B<sub>0</sub>)</b>				
Intercept (G <sub>00</sub> )	207.475	0.885	.000	--
PREAD07 (G <sub>01</sub> )	18.707	5.510	.002	1.083
PLUNCH (G <sub>02</sub> )	-23.833	11.548	.043	-1.380
TRT (G <sub>03</sub> )	1.739	1.246	.168	0.101
<b>Model for BASEISAT Slope (B<sub>1</sub>)</b>				
Intercept (G <sub>10</sub> )	0.451	0.032	.000	0.026
<b>Model for BASEMATH Slope (B<sub>2</sub>)</b>				
Intercept (G <sub>20</sub> )	0.160	0.032	.000	0.009
<b>Model for GENDER Slope (B<sub>3</sub>)</b>				
Intercept (G <sub>30</sub> )	-2.072	0.857	.016	-0.120
<b>Model for IEP Slope (B<sub>4</sub>)</b>				
Intercept (G <sub>40</sub> )	-9.958	1.080	.000	-0.576
<b>Model for LUNCH Slope (B<sub>5</sub>)</b>				
Intercept (G <sub>50</sub> )	5.077	2.242	.024	0.294
<b>Random Effects (Var. Components)</b>	<b>Variance</b>	<b>df</b>	<b>Chi-Square (p)</b>	<b>ICC</b>
Var. in school means (U <sub>0</sub> )	10.869	59	121.669 (.000)	0.055
Var. within schools (R)	187.422			

As these models show, several covariates had a significant impact on students' reading achievement. As shown in the final model (Table 14), among 6th-grade Tier 3 students, higher individual baseline reading and mathematics scores were found to predict higher reading achievement on the posttest. In addition, females and general education students had significantly higher posttest scores than males or special education students when holding all other variables constant. At the school level, a higher proportion of students reading at or above grade level was found to predict higher individual reading performance on the posttest, even among students whose individual baseline scores were below grade level. In addition, students in schools with larger proportions of low-income students were more likely to score lower on the spring 2008 ISAT, regardless of their own socioeconomic status. However, the full model (Table 13) shows that none of the interactions between treatment and NCLB groupings (including specific racial/ethnic groups such as Black and Hispanic, special education status [students with an individualized education plan], and low-income status) were significant. This means that although some of these characteristics were correlated with students' reading scores, there was no evidence of any differential program impact on any of these subgroups.

Results of analyses of 6th-grade Tier 2 students that explored variables representing the different subgroups are presented in Tables 15 and 16, below, including results for the full model and the final model, respectively.

**Table 15**  
**Full impact model for grade 6 Tier 2 students**

Fixed Effects	Coefficient	SE	<i>p</i>	Effect Size
Model for mean school reading achievement (B <sub>0</sub> )				
Intercept (G <sub>00</sub> )	224.215	1.172	.000	--
PMIN (G <sub>01</sub> )	-16.840	22.288	.453	-1.254
PREAD07 (G <sub>02</sub> )	4.275	9.830	.665	0.318
PSPED (G <sub>03</sub> )	10.457	21.184	.623	0.778
PLEP (G <sub>04</sub> )	-11.332	19.370	.561	-0.843
PLUNCH (G <sub>05</sub> )	-30.077	18.248	.105	-2.239
SIZE08 (G <sub>06</sub> )	-0.004	0.010	.645	0.000
COHORT (G <sub>07</sub> )	-0.960	2.277	.675	-0.071
TRT (G <sub>08</sub> )	2.226	1.729	.204	0.166
Model for BLACK Slope (B <sub>1</sub> )				
Intercept (G <sub>10</sub> )	-1.422	3.480	.683	-0.106
Model for HISPANIC Slope (B <sub>2</sub> )				
Intercept (G <sub>20</sub> )	-0.997	3.038	.743	-0.074
Model for BASEISAT Slope (B <sub>3</sub> )				
Intercept (G <sub>30</sub> )	0.317	0.049	.000	0.024
Model for BASEMATH Slope (B <sub>4</sub> )				
Intercept (G <sub>40</sub> )	0.155	0.035	.000	0.012
Model for GENDER Slope (B <sub>5</sub> )				
Intercept (G <sub>50</sub> )	0.205	1.131	.857	0.015
Model for IEP Slope (B <sub>6</sub> )				
Intercept (G <sub>60</sub> )	-5.194	2.312	.025	-0.387
Model for LUNCH Slope (B <sub>7</sub> )				
Intercept (G <sub>70</sub> )	1.843	2.735	.501	0.137
<b>Random Effects (Var. Components)</b>	<b>Variance</b>	<b>df</b>	<b>Chi-Square (<i>p</i>)</b>	<b>ICC</b>
Var. in school means (U <sub>0</sub> )	18.723	53	106.940 (.000)	0.121
Var. within schools (R)	136.181			

**Table 16**  
**Final impact model for grade 6 Tier 2 students**

Fixed Effects	Coefficient	SE	<i>p</i>	Effect Size
Model for mean school reading achievement (B <sub>0</sub> )				
Intercept (G <sub>00</sub> )	223.922	1.062	.000	--
PLUNCH (G <sub>01</sub> )	-39.817	12.735	.003	-2.964
TRT (G <sub>02</sub> )	2.690	1.540	.085	0.200
Model for BASEISAT Slope (B <sub>1</sub> )				
Intercept (G <sub>10</sub> )	0.312	0.047	.000	0.023
Model for BASEMATH Slope (B <sub>2</sub> )				
Intercept (G <sub>20</sub> )	0.160	0.033	.000	0.012
Model for IEP Slope (B <sub>3</sub> )				
Intercept (G <sub>30</sub> )	-4.918	2.235	.028	-0.366
<b>Random Effects (Var. Components)</b>	<b>Variance</b>	<b>df</b>	<b>Chi-Square (<i>p</i>)</b>	<b>ICC</b>
Var. in school means (U <sub>0</sub> )	16.329	59	112.309 (.000)	0.108
Var. within schools (R)	134.786			

As these results show, among 6th-grade Tier 2 students, an individual student's baseline reading and mathematics scores were both significant positive predictors of their spring 2008 reading achievement, while special education status was found to significantly predict lower scores. In addition, having a higher proportion of low-income students in a school was again found to predict lower reading achievement on the posttest, even among students who were not, themselves, economically disadvantaged.



## V. Evaluation of the School-Wide Impact of the Chicago Striving Readers Initiative in Years 1 and 2

### Study Design

#### Research Questions

The research questions relating to assessment of the overall impact of the Chicago Striving Readers initiative that were explored during the first two years of the evaluation are presented in Table 17, below. In Year 2, research questions were added to the study relating to exploring the initiative’s differential impact on students in different demographic subgroups.

**Table 17**  
**Research questions relating to the school-wide impact  
of the Chicago Striving Readers Initiative**

Year 1	Year 2
<b>Impact on Classroom Practices</b>	
1) Were the Chicago Striving Readers schools different than control schools on the five key components: professional development; comprehensive assessments; data-driven decision making; high-quality, high-interest materials; and comprehensive instruction?	1) Same as Year 1 [for seven revised key components].
<b>Impact on Students</b>	
1) What is the impact of the Chicago Striving Readers program on sixth through eighth grade students’ (i.e., whole-school impact) reading scores, as measured by ClassViews and the ISAT?	1) What is the impact of the Chicago Striving Readers program on sixth through eighth grade students’ (i.e., whole-school impact) reading scores, as measured by the ISAT? <sup>51</sup>
2) What is the [differential] impact of the Chicago Striving Readers program on [sixth vs. seventh vs. eighth] grade students’ reading scores (i.e., grade impact), as measured by ClassViews and ISAT?	2) What is the differential impact of the Chicago Striving Readers program on sixth vs. seventh vs. eighth grade students’ reading scores, as measured by the ISAT? <sup>42</sup>
--	3) Is there a differential impact of the Chicago Striving Readers program on the reading scores of sixth through eighth grade students in different NCLB subgroups, including gender, race, socioeconomic status (as determined by eligibility for free/reduced-price lunch), and ELL status, as measured by the ISAT? <sup>42</sup> ?

#### Sampling Plan

The process for selection of participating schools and random assignment to treatment and control groups was described under Section IV, as was the definition of the intent-to-treat population and the process for matching to students at control schools. For the purpose of assessing overall program impact school-wide, students at all three tiers are included in the analyses.

<sup>51</sup> As discussed above, ClassViews was not used as an outcome measure in Year 2 because it is no longer administered consistently.

### *Sample Size and Power*

The overall population of students with available data was described in Section IV. Power analyses for overall program impact were conducted using the same assumptions as those for the targeted and intensive intervention models. Specifically:

- Two-level HLM model (student and school)
- Type I error rate (alpha) = 0.05
- Intra-class correlation (rho) = 0.10
- Number of clusters (schools) = 63

The additional assumption regarding cluster size was as follows:

- Cluster size (n) = 148 (average number of intent-to-treat students per school with complete data)

The power to detect an effect size of at least .33, as well as the smallest effect size that could be detected at an ideal power level of 80%, were determined. Results of the power analyses are presented below.

#### Results:

Effect size = .333      Power = 98%

Effect size = .231      Power = 80% (ideal power level)

Power analysis results again show that the distribution of students and schools in this study provides sufficient statistical power to detect a program effect of less than one quarter standard deviation.

### *Data Collection Plan*

Assessment of the program impact on classroom practices involved administering the same data collection instruments to treatment and control schools in order to facilitate comparisons of instructional initiatives at the Striving Readers schools to what they might have been doing in lieu of the program, as inferred by the initiatives found to be in place at the control schools. Among the instruments described in Sections II and III that were used to collect data on classroom practices and program fidelity in the Striving Readers schools, the following were also administered at control schools during Year 2, using the same administration schedule:

- Classroom Observation Protocol
- Pre-Observation Literacy Environment Checklist
- School Leader Interviews
- Literacy Improvement Survey for Teachers (LIST)

For student outcomes, the same instrument—the ISAT—and the same data collection schedule were used for analyses of overall program impact on student achievement as previously described for the targeted and intensive intervention analyses in Section IV.

### *Summary of Analytic Approach to the Impact Analysis*

For analyses of the impacts on classroom practices, the original plan was to generate fidelity scales for the control and treatment schools and by comparing the scores for the two groups, to determine in which areas participating schools appear to have changed their practices the most, relative to what might have been expected in lieu of the program (*i.e.*, the “counterfactual”). However, response rates from control schools to the LIST survey were too low to provide an accurate representation of their practices. Because this survey was one of the primary sources of data for the fidelity scales, they could not be constructed for control schools. Nevertheless, direct comparisons of results from the classroom observations and pre-observation literacy environment checklist, as well as the school leader interviews, were still possible and provided useful information about what the control schools’ literacy initiatives imply about the counterfactual. Comparisons of observation codes from Striving Readers and control schools are discussed later in this section; discussions about the implications of school policies and classroom practices in the control schools that were obtained from school leader interviews are included in the discussion of qualitative findings about program implementation in Appendix D. In all cases where such comparisons are drawn however, it is important to note that evidence that focuses on the availability of resources, and/or the frequency of use of resources or instructional methods, does not necessarily reflect additional variations between Striving Readers and control schools resulting from differences in the specific content, nature and quality of these resources and methods. For example, as discussed below, control school classes were often observed to use “targeted materials” including media centers, listening centers and/or text sets. However, the texts included in the control schools’ text sets might not be of the same appropriateness or quality as those identified for Striving Readers schools, and control schools might not make as effective use of any of these materials for the purpose of providing differentiated instruction.

This section also presents the results from HLM analyses that were designed to assess the overall impact of the Striving Readers Initiative on students’ reading performance, as well as whether there were differential impacts of the program on different grade level and NCLB subgroups. These impact analyses again focused on the intent-to-treat populations that are described later in this section. Because school was the unit of assignment but impacts are measured at the student level, two-level models were again used for these cross-sectional analyses in order to account for the clustering of students in schools.

Two simpler models (full and final models) were fit to address the first research question. Additional analyses exploring the interactions between treatment and subgroups were also conducted to address research questions pertaining to differential impact by grade and by demographic group. Results of these latter analyses are presented under *Additional Analysis* later in this section.

Appendix E describes in greater detail the approaches used to fit these models, the specification of the models, the selection of covariates, and the treatment of missing data. An overall description of the sample used for these analyses is provided below.

## Description of the First- and Second-Year Samples

### *Number and Basic Characteristics of Schools, Teachers, Classrooms, and Students in First-Year Sample*

School characteristics were presented in Section IV. The demographic characteristics of all Striving Readers and control students, across all three tiers and across both cohorts, are summarized in Table 18. As these data show, the overall Striving Readers and control groups were demographically very similar, with the only exception being that Cohort 1 control schools had a somewhat higher representation of African American students and a somewhat lower representation of Hispanic students than Striving Readers schools. The difference between cohorts in the proportion of African American students that was seen among Tier 2 and 3 students above is also apparent for the schools as a whole.

**Table 18**  
**Characteristics of intent-to-treat students with complete data:\* Tiers 1–3**

Characteristics		All Students		Cohort 1 Students		Cohort 2 Students	
		Control (N=4,688)	Treatment (N=4,655)	Control (N=2,655)	Treatment (N=2,642)	Control (N=2,033)	Treatment (N=2,013)
Grade level	6	31%	31%	30%	29%	33%	33%
	7	32%	32%	33%	34%	32%	29%
	8	37%	37%	38%	37%	36%	38%
Gender	Female	50%	49%	49%	50%	50%	49%
	Male	50%	51%	51%	50%	50%	51%
Race/Ethnicity	American Indian	0%	0%	0%	0%	0%	0%
	Asian	0%	1%	1%	2%	0%	0%
	African American	72%	58%	55%	29%	94%	97%
	Hispanic	23%	34%	37%	59%	5%	2%
	White	3%	3%	5%	6%	0%	0%
	Other/Multiracial	2%	3%	3%	4%	1%	1%
% Special education		14%	14%	12%	13%	16%	16%
% ELL		1%	1%	1%	2%	1%	0%
% Free/reduced-price lunch eligible		92%	93%	89%	91%	97%	95%

\*Baseline scores, tier assignments, and outcome data were available for all students, making it possible to include them in impact analyses.

### **Impacts on Teacher Practices and Classroom Instructional Environment at the End of Two Years**

Data were available from spring 2008 classroom observations<sup>52</sup>—including from the pre-observation checklists as well as the observation codes themselves—that provided evidence about impact for several key program components. These data provided evidence about program

<sup>52</sup> Observation data from the 2007–2008 school year are being analyzed only from spring observations, because these were the only observations conducted by Metis Associates.

components relevant to the whole-school/blended intervention model, as well as some that are specific to the targeted intervention.<sup>53</sup> Data were available from a total of 47 6th-grade English language arts classes at treatment schools and 54 classes at control schools that were observed during spring 2008. Wherever possible, these data are disaggregated by cohort, and are also compared to analogous observation results from the first program year as reported in the Year 1 Implementation Report. It is expected that implementation of the Striving Readers initiative would result in an increase in the availability and use of resources, as well as in changes in the way in which the resources are applied and in the use of new and innovative instructional methods and strategies. While not proving a causal relationship, any differences in this regard that are observed between treatment and control schools can therefore be taken as supporting evidence of the impact of the program on classroom practices.<sup>54</sup>

### *Use of High-Quality, High-Interest Materials*

This program component focuses on the use of highly motivating reading materials integrated with engaging technology and audio resources. One element of this component is the availability of media centers—with at least three computers and one printer per classroom—that are made accessible to students through the Striving Readers grant for individual and small-group work. Spring observations revealed that treatment classrooms were clearly better equipped in this regard than control school classrooms: almost all of the observed treatment classrooms had media centers meeting these standards, compared to fewer than half of control school classrooms (88% vs. 45%). On average, treatment classrooms also had one more computer per media center than did controls (four computers vs. three), as well as a larger number of printers per media center (averaging 1.3 printers vs. 0.98).

Through grant support, Striving Readers classrooms are also equipped with listening centers—learning centers where students can access models of fluency, record themselves while reading, monitor their own fluency levels, and listen to audio books. Again, the majority of treatment classrooms (78%) were observed to have listening centers that were in working order, compared with just over half (56%) of control school classes. In addition, three quarters (75%) of treatment classrooms had several sets of headphones and almost two thirds (63%) had audio materials for students’ use, which was substantially more than were observed in control schools (where 52% and 43% of classrooms, respectively, had these resources).

Striving Readers and control classrooms were similar in terms of the availability of libraries that were easily accessible to students: in both groups, almost all observed classes (97% and 95%, respectively) had such libraries. The groups were also comparable in terms of the content, organization, and quality of their libraries: nearly all libraries observed had both fiction and nonfiction books available (95% of libraries in both groups), were organized and in good shape (93% of libraries from both groups), and had a variety of texts to appeal to readers of differing abilities and interests (93% of treatment libraries and 89% of controls). In most cases, books were grouped by genre (85% of treatment libraries and 84% of controls), were clearly labeled (82% and 78%, respectively), and had a checkout system in place (76% of classrooms from both

---

<sup>53</sup> Most of the AMP classes had already ended by the time spring observations were conducted, resulting in insufficient data to provide evidence about the intensive intervention.

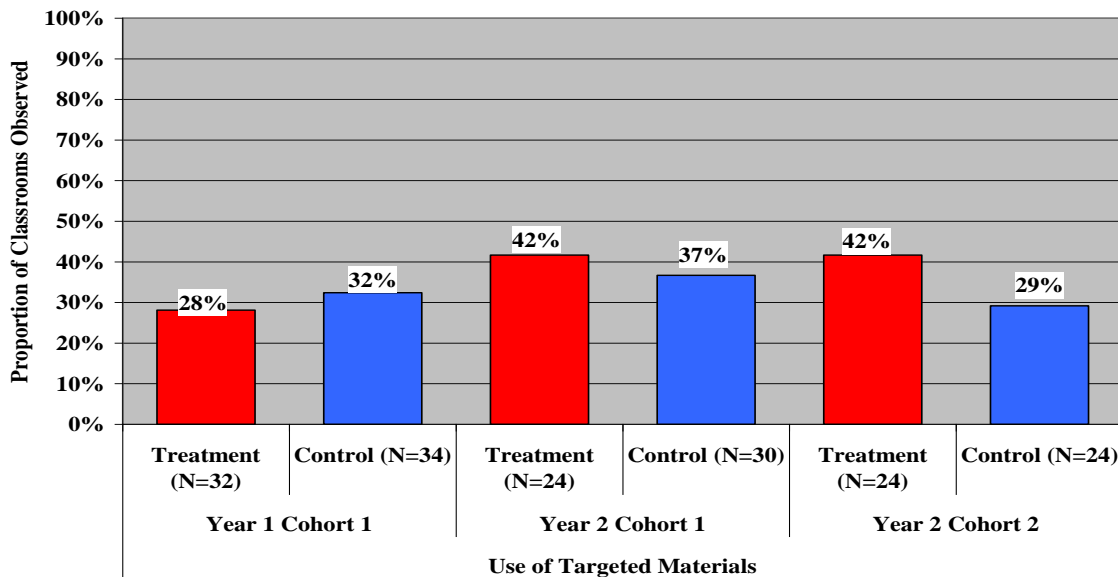
<sup>54</sup> It should be noted that the observation results reported below provide evidence about the availability and use of resources and instructional methods, but generally do not address the quality of their use.

groups). Most classrooms from both groups were also observed to have text sets (84% and 89% of classrooms, respectively). Newspapers and magazines were observed considerably less frequently in both groups: fewer than half of classrooms (45% of Striving Readers classrooms and 41% of controls) had magazines, while newspapers were observed the most infrequently, in only 21% and 29% of classrooms, respectively.

Although the fledgling program for adopting the use of handheld computers (palm pilots) to support literacy instruction was only in its preliminary, pilot stages in the second program year, 15% of Striving Readers classrooms (but none of the control schools) had handheld computers available in spring 2008.

An overall summary of the proportions of classrooms that were observed to be actually *using* any of the primary targeted materials (listening centers, computer media centers, or text sets) during a literacy lesson are presented by year and cohort in Figure 1, below.

**Figure 1**  
**Spring 2007 and spring 2008 classroom observations**  
**Variation in the use of targeted materials by year and cohort**



As shown in Figure 1, during Year 1, slightly more control classrooms than Striving Readers classrooms (32% vs. 28%) were observed using these targeted materials. However, this trend reversed in Year 2, when greater proportions of treatment classrooms from both cohorts implemented targeted materials during the observed lessons (42% vs. 37% for Cohort 1 and 42% vs. 29% for Cohort 2).

### *Use of Core Comprehension Strategies*

A major focus of the Striving Readers program is to make comprehension the core of all literacy instruction, and specifically to increase the focus on the following seven comprehension strategies: summarizing, predicting, questioning, inferring, text structure, visualizing, and metacognition. Observers coded whether any of these strategies were included among instructional activities during an observed literacy lesson. Figure 2 shows the proportion of classrooms that exhibited at least one of these seven strategies.

**Figure 2**  
**Spring 2007 and spring 2008 classroom observations**  
**Variation in the use of comprehension strategies by year and cohort**

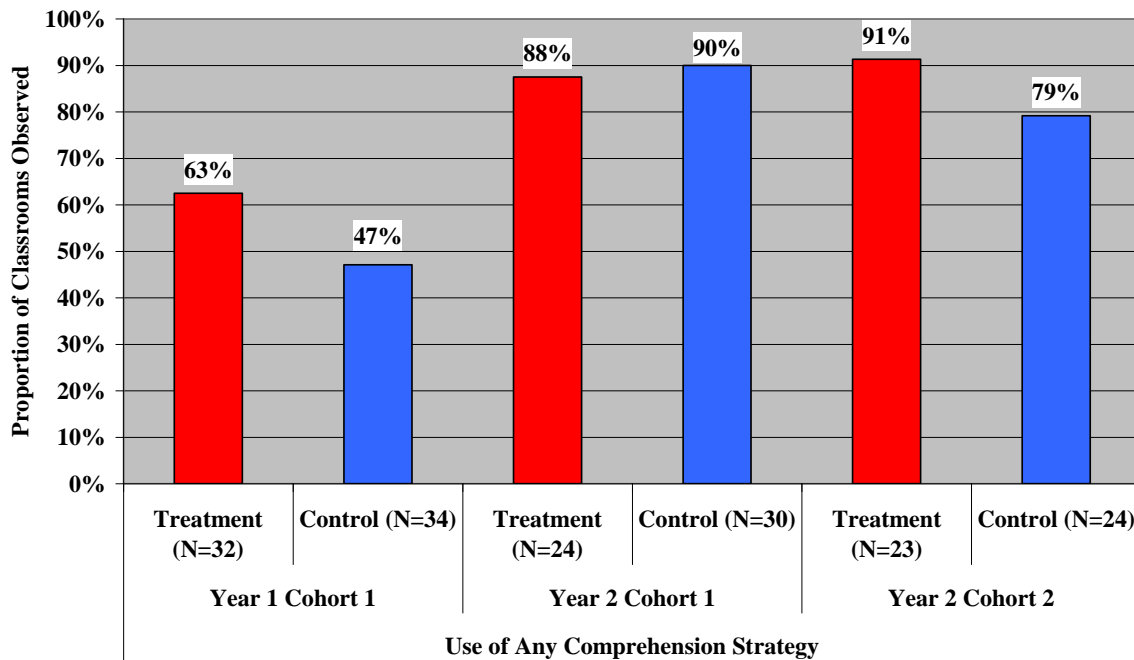
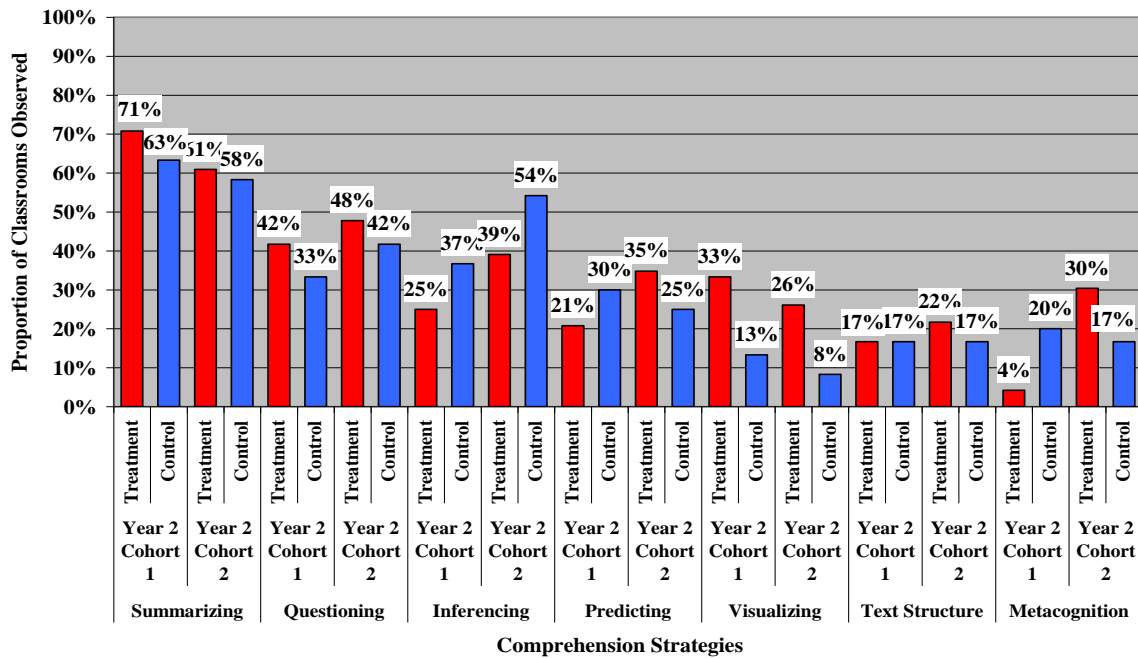


Figure 2 shows that comprehension strategies were used more frequently in Striving Readers classrooms than in control schools in both years, but that both groups showed a marked increase in the frequency of use of comprehension strategies by Year 2. Although the difference between treatment and control schools was less pronounced in the second year, this appears to be the result of increased use in control schools combined with ceiling effects in the Striving Readers schools, where the vast majority of classrooms were already using these strategies. Specifically, in Year 1, more than three fifths (63%) of treatment classrooms covered at least one comprehension strategy during the observation, compared with less than half (47%) of the control school classrooms. In Year 2, the vast majority of both treatment and control classrooms in Cohort 1 schools (88% and 90%, respectively) were observed to use these strategies; among Cohort 2 schools, usage was more universal among treatment classes (91% of observed lessons) than among controls (where 79% used these strategies).

An analysis of the specific strategies that were used is also revealing, although such an analysis was not available for Year 1. Figure 3 and Table 19, below, display the frequency with which specific comprehension strategies were observed during literacy lessons during spring 2008 observations.



**Figure 3**  
**Spring 2008 classroom observations**  
**Comparison of classrooms' use of comprehension strategies by cohort**



**Table 19**  
**Spring 2008 classroom observation**  
**Use of specific comprehension strategies by cohort**

Cohort	Group	Mean Proportion						
		Summarizing	Questioning	Inferring	Predicting	Visualizing	Text Structure	Metacognition
Cohort 1	Treatment	70.8%	41.7%	25.0%	20.8%	33.3%	16.7%	4.2%
	Control	63.3%	33.3%	36.7%	30.0%	13.3%	16.7%	20.0%
Cohort 2	Treatment	60.9%	47.8%	39.1%	34.8%	26.1%	21.7%	30.4%
	Control	58.3%	41.7%	54.2%	25.0%	8.3%	16.7%	16.7%

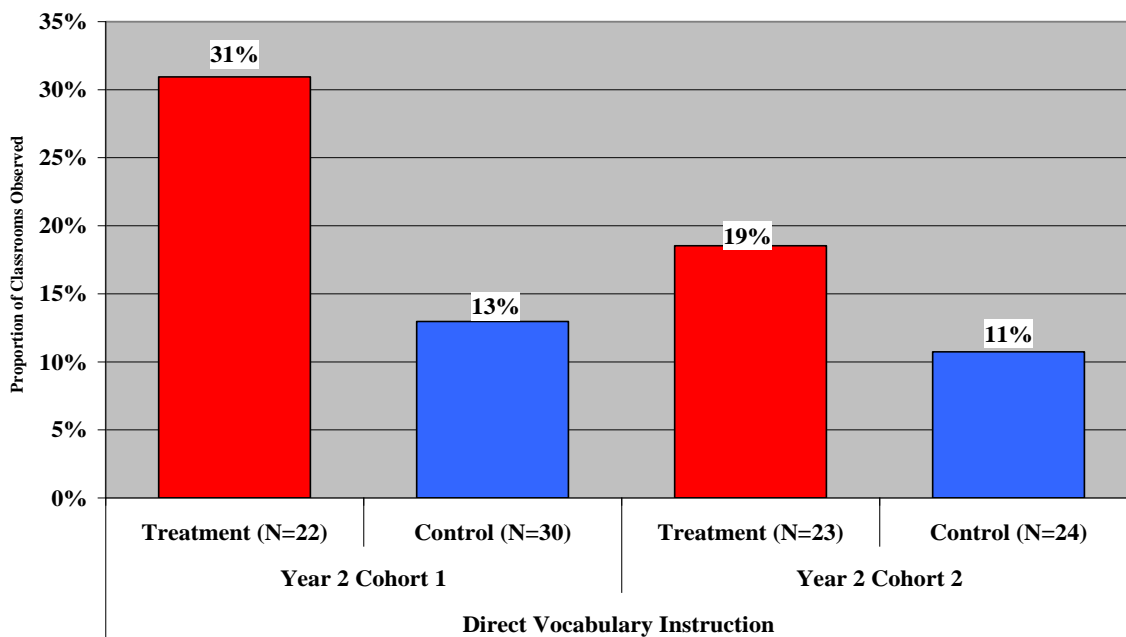
As the above data show, *summarizing* was by far the most frequently used comprehension strategy during these literacy classes, with more than two thirds of Cohort 1 treatment classes and about three fifths of Cohort 1 control classes and Cohort 2 treatment and control classes using this strategy. For most comprehension strategies, Striving Readers classes were observed conducting instructional activities that used the strategy more frequently than control classrooms. This difference was most dramatic for frequency of use of *visualizing* in both cohorts as well as use of *metacognition* among Cohort 2, where in all cases the frequency of use in Striving Readers classrooms ranged from almost double to more than triple that of controls. For a few strategies, however, the trend was reversed, with control classes using the strategy more

frequently. This was true in both cohorts for *inferring*, and in Cohort 1 for *predicting* and *metacognition*.

### *Use of Vocabulary Instruction*

The Striving Readers program also promotes a focus on direct vocabulary instruction. Observation data were analyzed to determine the proportion of treatment and control classrooms that exhibited direct instruction in this area during an observed lesson. Figure 4 displays these results.

**Figure 4**  
**Spring 2008 classroom observations**  
**Variation in the use of direct vocabulary instruction by cohort\***



\*Two Cohort 1 treatment school observations were excluded from this analysis because the observation checklist was coded improperly.

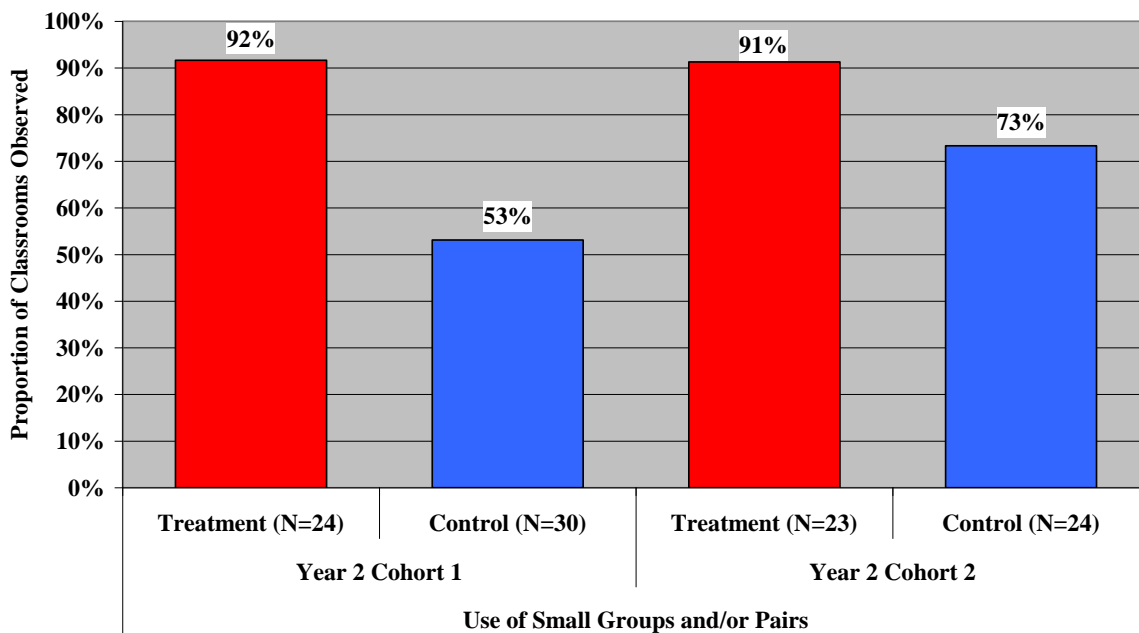
As shown in Figure 4, substantially greater proportions of treatment classrooms than control classrooms from both cohorts demonstrated direct vocabulary instruction during their observed lessons, and among treatment classrooms, direct vocabulary instruction was observed more frequently among Cohort 1 schools than Cohort 2. Specifically, among Cohort 1 schools, 31% of treatment classrooms were observed to use this technique, compared with 13% of control classrooms. Within Cohort 2, the observed frequencies of use were 19% and 11% for treatment and control schools, respectively.

### *Use of Small-Group Instruction*

Another major focus of the Striving Readers Initiative is to promote the use of small-group activities as a means of differentiating instruction for all students. Ideally, this should take place within the context of whole-part-whole instruction, where the small-group instruction is preceded and followed by whole-group activities that provide context and an opportunity to synthesize what was learned. To ensure continuity of instruction, another critical characteristic of small-group instruction is that it should be aligned with other work that the rest of the class is doing and that the students in the small group are doing at other times.

In the Year 1 report, 53% of the treatment classrooms and 21% of the control classrooms were reported to be using whole-part-whole instruction. However, due to ambiguities in the observation checklist that was used to code the classroom observations in both years, Metis Associates felt that too many assumptions were needed in order to interpret the observation data as to whether whole-part-whole instruction had actually taken place during the observed lesson. Instead, the data for Year 2 were analyzed at a simpler but more reliable level for the proportion of classroom observations where instruction was provided in small groups and/or pairs.<sup>55</sup> These data are displayed in Figure 5.

**Figure 5**  
**Spring 2008 classroom observations**  
**Variation in the use of small groups and/or pairs**



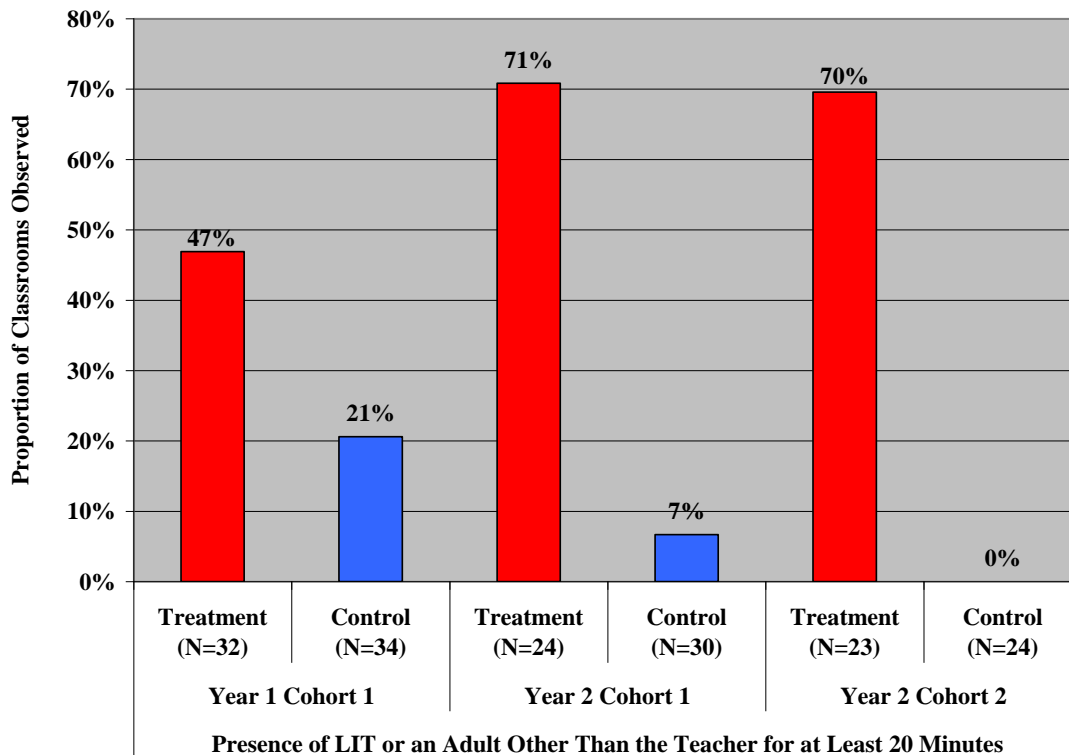
<sup>55</sup> Data collection instruments for Year 3 are being reviewed so that the use of whole-part-whole instruction can be captured more accurately.

Figure 5 shows that a substantially greater proportion of treatment classrooms were observed using small groups and/or pairs than control classrooms, especially among Cohort 1 classrooms. Nearly all of the Cohort 1 and 2 treatment classrooms (92% and 91%, respectively) grouped students into small groups and/or pairs for literacy instruction, as compared to half (53%) of the Cohort 1 control and about three quarters (73%) of the Cohort 2 control classrooms.

**Targeted Intervention: Presence of the LIT**

Classroom observation data were analyzed to determine whether or not an LIT was present in the observed Striving Readers classrooms for at least 20 minutes.<sup>56</sup> For comparison, control school classes were coded for the presence of an adult other than the classroom teacher for at least 20 minutes. For this purpose, “any other adult” was defined by the use of the codes for reading specialist, other special teacher, an aide, or librarian. These data, along with what was reported for spring 2007, are presented in Figure 6.

**Figure 6**  
**Spring 2007 and spring 2008 classroom observations**  
**Variation in the presence of an LIT (treatment) or other adult (control) by year and cohort**



<sup>56</sup> Although the LIT is supposed to be present in the classroom for a full hour, the criterion of a minimum of 20 minutes—the amount of time that they are supposed to spend providing scaffolded instruction in a small group setting to struggling readers—was used for these analyses to make them comparable to the results reported for Year 1.

As shown in Figure 6, a much greater proportion of treatment classrooms had an LIT present for at least 20 minutes during the observations in both years, and in both cohorts in Year 2. In addition, the proportion of Striving Readers classes where an LIT was present increased substantially from Year 1 to Year 2.<sup>57</sup> In spring 2007, only one in five (21%) Cohort 1 control school classrooms had an adult other than the classroom teacher present in the classroom for at least 20 minutes, compared with about half (47%) of the Cohort 1 treatment school classrooms. In spring 2008, an adult other than the classroom teacher was observed in only 7% of the Cohort 1 control school classes and in none of the Cohort 2 control school classes. In comparison, the majority of Cohort 1 and 2 treatment school classes had an LIT present for 20 minutes or more (71% and 70%, respectively).

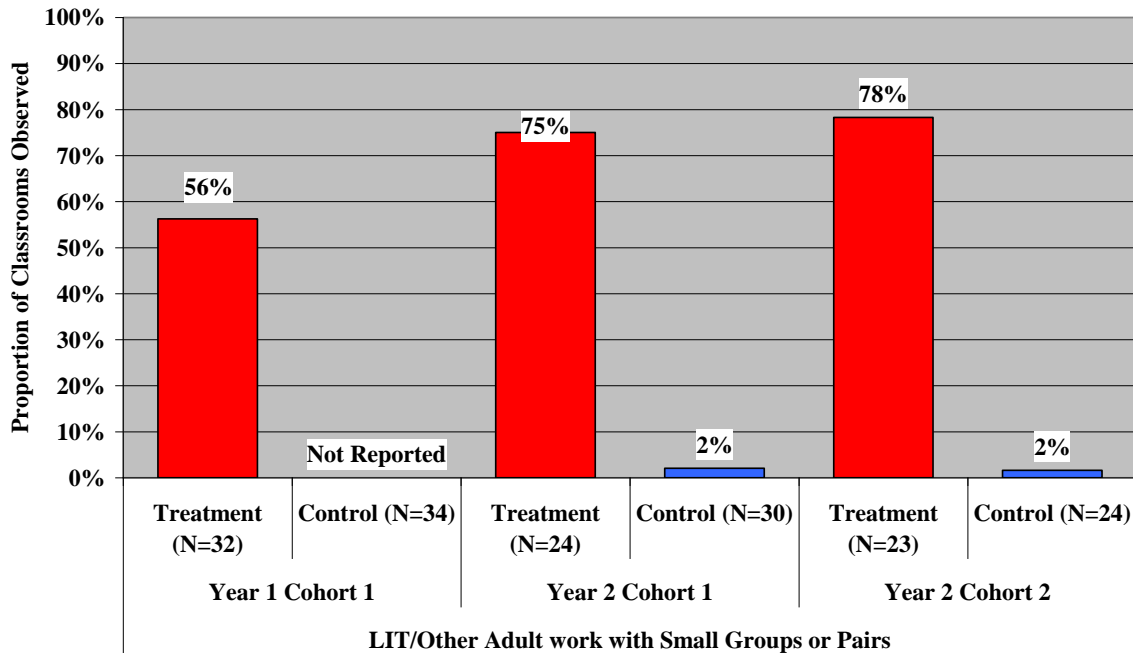
### ***Targeted Intervention: Small-Group Instruction Provided by the LIT***

One of the major roles that the LIT plays within the regular literacy classroom is to provide targeted, small-group instruction for struggling readers. Beyond the mere presence of the LIT in the classroom, it is this function that defines the targeted intervention model. Figure 7 presents the proportion of classrooms where an LIT or, in the case of the control schools, an adult other than the classroom teacher, was observed to provide instruction to a small group and/or pair of students.

---

<sup>57</sup> It is possible, however, that the actual proportion of classes served by an LIT was higher than reported for Year 1: since literacy blocks generally last from 90 to 120 minutes and the LIT is only expected to be in the classroom for 60 minutes, it is possible that some observations missed the LIT if they did not stay for the entire period. This would not be true for the spring observations in Year 2 however, since all observations lasted for the entire duration of the literacy block.

**Figure 7**  
**Spring 2007 and spring 2008 classroom observations**  
**Variation in the LIT (treatment) or other adult (control) providing instruction to small groups/pairs by cohort**



As Figure 7 shows, it appears from the classroom observation data that while some control classes did have a second adult present, this particular role was virtually unique to the LIT in the Striving Readers class. Although comparisons to control schools cannot be made for Year 1 (when these data were not reported for controls), during Year 2 an adult was observed providing small-group instruction in only 2% of the Cohort 1 and Cohort 2 control school literacy classes, compared with three fourths of treatment classes (75% and 78% of Cohort 1 and Cohort 2 classes, respectively). Moreover, the proportion of classes where an LIT was observed providing small-group instruction increased substantially compared with the spring of the prior year, when an LIT was observed providing small-group instruction in just over half (56%) of observed Striving Readers classes.

***Intensive Intervention: After-School Literacy Program Provided by the LIT***

Because the Year 2 classroom observations conducted by Metis Associates took place in the spring of the 2007-08 school year after most after-school programs had ended, it was not possible to obtain detailed evidence comparing the nature of intervention that Tier 3 students received in the AMP class to that of any after-school programs that control schools may have conducted. All 31 Striving Readers schools implemented an AMP after school program that averaged approximately 240 minutes per week for approximately 26 weeks. Most of the control schools (86% of Cohort 1 schools and 87% of Cohort 2 schools) also reported offering after-school programming specifically targeting struggling readers. These schools implemented a

variety of after-school programs for these students. While they did not typically use the AMP program, details about content, scheduling and attendance for these programs were not obtained.

### **Discussion and Conclusions**

Data for the comparison of classroom practices in Striving Readers versus control schools was unfortunately somewhat limited during the second program year. However, a clear trend emerged from the observation data that indicates that although the availability of resources was not dramatically different, more of the instructional methods of the Striving Readers Initiative were present in Year 2 than in Year 1, and in many cases, were also more prevalent in Striving Readers classrooms than in control classrooms. These results imply not only that implementation of the Striving Readers Initiative is progressing in a positive direction, but also that the instructional activities that schools are using are not merely the result of larger educational trends in Chicago, but are specifically attributable to the program.

It must be cautioned, however, that there are limitations to the interpretation of these observation data. Because observers are in the classroom for only the briefest period of time during the course of the year, the proportion of observations during which particular activities are observed may not be representative of the level of activity throughout the year, and it is impossible to know whether the observations might be over- or under-representing the frequency of observed activities. Nevertheless, there is no *a priori* reason to expect that any type of activity would be more over- or under-represented at treatment or control schools. Thus, while the absolute frequency of observed activities may not always be representative, relative comparisons between Striving Readers and control schools can be expected to be more reliable.

Another limitation of observation data is that there was not a sufficient number of observations to be able to draw conclusions about activity at the individual school level. For this reason, these observation data cannot provide insight into the variation in program implementation or impact across schools.

### **Impacts on Students**

Table 20 shows the estimates of the overall impacts of the Chicago Striving Readers program on all intent-to-treat students in grades 6–8 and in Tiers 1–3.<sup>58</sup>

**Table 20**  
**Estimated overall impact of Striving Readers on all students**

Population Group	Unadjusted Means		Regression-Adjusted Means		Estimated Impact	Effect Size	p Value
	Control	Treatment	Control	Treatment			
All students	232.35	235.47	233.23	234.37	1.14	0.05	.099

<sup>58</sup> Impact estimates are based on the final simple model that did not include interaction terms. This model is presented in Table F-9 in Appendix F.

As shown in Table 20, the analyses did not reveal any significant overall impact of the Striving Readers Initiative on students' reading performance, as measured by their spring 2008 ISAT reading scale scores. Specifically, students in the treatment group had slightly higher mean reading scale scores than students in the control group, but these differences were not significant and the effect size was very small (.05). However, further examination of the covariates in the model provides a better understanding of the differential program impacts for various demographic sub-groups. These results are discussed in the next section.

### **Additional Analyses**

Additional analyses were conducted to assess what other factors might also contribute to reading achievement, and whether there was a differential impact of the Striving Readers Initiative for students in different grades and in different NCLB subgroups. For these analyses, grade level and NCLB subgroup variables, among others, were entered into the model as covariates. In addition, the model also explored the interaction effects between these subgroups and treatment to determine whether treatment effects were greater for some groups than others. Results are presented in Tables 21 and 22 below, which present results for the full model and for the final model (after removing covariates with  $p$  values greater than .20), respectively.



**Table 21**  
**Full model for overall program impact with interactions**

<b>Fixed Effects</b>	<b>Coefficient</b>	<b>SE</b>	<b>p</b>	<b>Effect Size</b>
<b>Model for mean school reading achievement (B<sub>0</sub>)</b>				
Intercept (G <sub>00</sub> )	233.266	0.492	.000	--
PMIN (G <sub>01</sub> )	5.792	9.032	.524	0.252
PREAD07 (G <sub>02</sub> )	6.507	3.921	.102	0.283
PSPED (G <sub>03</sub> )	1.013	8.193	.903	0.044
PLEP (G <sub>04</sub> )	8.457	5.268	.114	0.368
PLUNCH (G <sub>05</sub> )	-11.084	6.965	.117	-0.482
SIZE08 (G <sub>06</sub> )	-0.007	0.004	.085	0.000
COHORT (G <sub>07</sub> )	-5.305	0.956	.000	-0.231
TRT (G <sub>08</sub> )	1.084	0.715	.135	0.047
<b>Model for BLACK Slope (B<sub>1</sub>)</b>				
Intercept (G <sub>10</sub> )	-2.370	1.036	.022	-0.103
TRT (G <sub>11</sub> )	1.947	1.402	.165	0.085
<b>Model for HISPANIC Slope (B<sub>2</sub>)</b>				
Intercept (G <sub>20</sub> )	-1.148	0.948	.227	-0.050
TRT (G <sub>21</sub> )	1.754	1.260	.164	0.076
<b>Model for GRD7 Slope (B<sub>3</sub>)</b>				
Intercept (G <sub>30</sub> )	-0.018	0.492	.971	-0.001
TRT (G <sub>31</sub> )	-0.522	0.704	.459	-0.023
<b>Model for GRD8 Slope (B<sub>4</sub>)</b>				
Intercept (G <sub>40</sub> )	2.865	0.514	.000	0.125
TRT (G <sub>41</sub> )	-3.163	0.736	.000	-0.138
<b>Model for BASEISAT Slope (B<sub>5</sub>)</b>				
Intercept (G <sub>50</sub> )	0.412	0.014	.000	0.018
TRT (G <sub>51</sub> )	0.026	0.020	.194	0.001
<b>Model for BASEMATH Slope (B<sub>6</sub>)</b>				
Intercept (G <sub>60</sub> )	0.195	0.012	.000	0.009
TRT (G <sub>61</sub> )	0.001	0.017	.940	0.000
<b>Model for TARGETED Slope (B<sub>7</sub>)</b>				
Intercept (G <sub>70</sub> )	-2.839	0.605	.000	-0.123
TRT (G <sub>71</sub> )	-0.825	0.852	.333	-0.036
<b>Model for INTENSIVE Slope (B<sub>8</sub>)</b>				
Intercept (G <sub>80</sub> )	-3.772	0.623	.000	-0.164
TRT (G <sub>81</sub> )	1.246	0.888	.161	0.054
<b>Model for GENDER Slope (B<sub>9</sub>)</b>				
Intercept (G <sub>90</sub> )	-1.430	0.385	.000	-0.062
TRT (G <sub>91</sub> )	-0.162	0.545	.767	-0.007
<b>Model for IEP Slope (B<sub>10</sub>)</b>				
Intercept (G <sub>100</sub> )	-8.251	0.626	.000	-0.359
TRT (G <sub>101</sub> )	1.333	0.891	.134	0.058
<b>Model for LUNCH Slope (B<sub>11</sub>)</b>				
Intercept (G <sub>110</sub> )	-0.629	0.729	.389	-0.027
TRT (G <sub>111</sub> )	-0.712	1.050	.498	-0.031
<b>Random Effects (Var. Components)</b>	<b>Variance</b>	<b>df</b>	<b>Chi-Square (p)</b>	<b>ICC</b>
Var. in school means (U <sub>0</sub> )	5.838	54	350.211 (.000)	0.034
Var. within schools (R)	165.526			

**Table 22**  
**Final model for overall program impact with interactions**

<b>Fixed Effects</b>	<b>Coefficient</b>	<b>SE</b>	<b>p</b>	<b>Effect Size</b>
<b>Model for mean school reading achievement (B<sub>0</sub>)</b>				
Intercept (G <sub>00</sub> )	233.223	0.473	.000	--
PREAD07 (G <sub>01</sub> )	6.161	3.725	.103	0.268
PLEP (G <sub>02</sub> )	7.406	4.898	.136	0.322
PLUNCH (G <sub>03</sub> )	-9.984	6.275	.117	-0.434
SIZE08 (G <sub>04</sub> )	-0.008	0.004	.025	0.000
COHORT (G <sub>05</sub> )	-5.295	0.934	.000	-0.230
TRT (G <sub>06</sub> )	1.144	0.678	.097	0.050
<b>Model for BLACK Slope (B<sub>1</sub>)</b>				
Intercept (G <sub>10</sub> )	-1.201	0.587	.041	-0.052
<b>Model for GRD8 Slope (B<sub>2</sub>)</b>				
Intercept (G <sub>20</sub> )	2.773	0.418	.000	0.121
TRT (G <sub>21</sub> )	-2.663	0.583	.000	-0.116
<b>Model for BASEISAT Slope (B<sub>3</sub>)</b>				
Intercept (G <sub>30</sub> )	0.414	0.012	.000	0.018
TRT (G <sub>31</sub> )	0.020	0.013	.123	0.001
<b>Model for BASEMATH Slope (B<sub>4</sub>)</b>				
Intercept (G <sub>40</sub> )	0.195	0.008	.000	0.008
<b>Model for TARGETED Slope (B<sub>5</sub>)</b>				
Intercept (G <sub>50</sub> )	-3.330	0.421	.000	-0.145
<b>Model for INTENSIVE Slope (B<sub>6</sub>)</b>				
Intercept (G <sub>60</sub> )	-3.136	0.444	.000	-0.136
<b>Model for GENDER Slope (B<sub>7</sub>)</b>				
x Intercept (G <sub>70</sub> )	-1.507	0.273	.000	-0.066
<b>Model for IEP Slope (B<sub>8</sub>)</b>				
Intercept (G <sub>80</sub> )	-8.284	0.617	.000	-0.360
TRT (G <sub>81</sub> )	1.342	0.872	.124	0.058
<b>Model for LUNCH Slope (B<sub>9</sub>)</b>				
Intercept (G <sub>90</sub> )	-0.976	0.523	.062	-0.042
<b>Random Effects</b>	<b>Variance</b>	<b>df</b>	<b>Chi-Square</b>	<b>ICC</b>
<b>(Var. Components)</b>			<b>(p)</b>	
Var. in school means (U <sub>0</sub> )	5.605	56	353.128 (.000)	0.033
Var. within schools (R)	165.500			

As shown in Table 21 (full model), at the student level, being African American, a male, or a student in special education were all significant negative predictors of students' reading performance; in other words, after holding all other covariates constant, students in one of these groups tended to have lower spring 2008 reading test scores than students who did not belong to these groups. Furthermore, students in grades 6 and 7 and students in Tiers 2 and 3 were also more likely to score lower than their peers after holding all other covariates constant. The first effect is to be expected; because the ISAT is vertically scaled, students in higher grades tend to obtain higher scores. The pattern seen for tiers is also not surprising, because these students were selected for the targeted and intensive interventions because of their lower reading skills. Students' individual baseline test scores in reading and mathematics were both positive predictors of their spring 2008 reading performance.

At the school level, when holding all other covariates constant, students in larger schools and students in Cohort 2 schools were more likely to score lower on the spring 2008 ISAT reading test than their peers.

As shown in Table 22 (final model), there was found to be a statistically significant, negative interaction between grade 8 and treatment. Although the effect size was small, these results suggest that the Striving Readers program has been more effective with students in the lower grades than with students in 8th grade. This also is not surprising; although resources and professional development were provided to all three grades, program implementation did not begin to focus on the 8th grade until Year 3. On the other hand, district staff *did* observe during Year 2 that achievement improved among 7th grade students from Cohort 1—the first group of students to have received more than one year of program intervention. This effect was not detected in these analyses, but since the models did not explore program impacts by grade and by cohort simultaneously, it is possible that the only reason it was not detected was that Cohort 2 7th graders did not do as well as Cohort 1 and therefore reduced the average gains for that grade when measured across cohorts.<sup>59</sup> None of the other interaction terms between subgroups of students and treatment were found to be statistically significant, revealing that the program has not had a differential impact on any of the subgroups that were included in the analyses, including subgroups based on students' gender, race, and special education status.

---

<sup>59</sup> Beginning in Year 3, if feasible, HLM analyses for overall program impact will explore the possibility of such effects by including a term for a three-way interaction between treatment, grade and cohort.

## VI. Summary and Conclusions

---

As might be expected in the second year of a five-year initiative, program implementation in Chicago’s Striving Readers initiative was found to occur at a medium level overall, and evidence exists that the program appears to be on a positive trajectory for further improvement. Further details of the implementation status and their implications for the analyses of student impacts are discussed below.

The initiative has struggled with the tier assignment process in its early years. In the first year, errors in the application of cut-off scores resulted in offering the intensive intervention to all students below grade level—a major modification of the intended model that would confound interpretation of impact analyses. (In addition, to the extent that the BRI was used to assign some borderline students to tiers in Year 1, it will cause further complications with interpretation of longitudinal analyses that will be conducted beginning in Year 3, since this assignment method cannot be replicated with control students for the purpose of creating matched samples. This does not affect the current analyses of Year 2 outcomes, however.) In Year 2 at least, because state IMAGE data were available in time for the tier assignment process, the BRI did not need to be used, increasing the validity of the process of matching students to their peers in control schools.<sup>60</sup>

Some program components had particularly strong fidelity ratings this year—most notably, the use of the whole-part-whole model and the gradual release model. However, because both of these ratings are based on self-report data, there remains a possibility that the ratings are at least somewhat inflated due to staff’s incomplete understanding of the models that they are rating and lack of awareness of what they might need to do differently. Indeed, program leaders have expressed concerns about this very issue during district staff interviews and ongoing communications with the evaluator. This is a common phenomenon with self-report data, one which can—not infrequently—occur to such a degree that self ratings *decline* over time when in fact the quality of implementation is improving. Nevertheless, the data are valuable if interpreted carefully: they can be as important for highlighting what participants do or do not know about the program as they are for indicating fidelity levels. Additional exploration that is being considered of the extent to which fidelity ratings are aligned with program implementation rubrics and the impressions of project leadership would help to illuminate these questions.

Although evidence about changes in program fidelity over the first two years is incomplete, and data that do exist are difficult to interpret because of the transition between evaluators, there was considerable evidence of notable progress from Year 1 to Year 2. A regular schedule of communications and professional development opportunities for all staff levels has proven particularly valuable for moving the initiative forward and strengthening the skills of key staff such as the Coordinators and the LITs. The “community of learners” that has grown out of this structure helps to ensure that staff are positioned to incorporate new concepts and to grapple with new challenges. The Coordinators also cite the commitment made to the Striving Readers

---

<sup>60</sup> Nevertheless, because these students did not have ISAT scores, they still could not be included in the analyses. Students who were missing both ISAT and IMAGE scores were also assigned based on their BRI scores, but again have not been included in analyses.

initiative by teachers and school administrators, and have noted the “budding” of best practices resulting from increasing confidence and initiative among teachers.

Among specific instructional initiatives, the available evidence does imply that substantial improvements may have been achieved in the use of the whole-part-whole model—albeit subject to the caveats about the fidelity scales mentioned above. However, other challenges that were encountered in Year 1 remained a challenge in Year 2. This included instructional focus on comprehension strategies, which appears to have been lacking in both years for at least some of the specific core strategies, and providing increased instructional time for Tier 3 students. In addition, even among some of the program components for which implementation was rated strongly on average, in some cases there was considerable variation among schools, with some schools demonstrating much less success in implementation. Program components with a high degree of variation among schools included the level of collaboration with the LIT, the amount of time allocated to the AMP after-school program, vocabulary instruction, and, even though substantial improvements were observed from Year 1 overall, in use of the whole-part-whole model.

Project leadership also acknowledged the conclusions that emerged from these variations in fidelity ratings. The Project Director observed that understanding the role of an “interventionist” in the classroom, as distinct from that of a teacher or traditional teacher’s assistant, posed a major challenge for a number of LITs. She concurred that the increase in small group instruction has been a “tremendous” success, but that as it was being applied in Year 2, this initiative still represented primarily a change in grouping, but not necessarily the implementation of differentiation. Taking this instructional strategy to the next level remained a challenge.

In summary, while not unexpected after only the second year of a five year initiative, by the end of the 2007-2008 school year there still remained considerable room for improvement. There was a consensus that further training and support were still needed at all levels of program implementation—from teachers and principals to LITs and even District Coordinators—and for the most part, each of these levels of staff also recognized their own need for further training. Interviews with project leadership painted a picture of a comprehensive, complex and demanding program, and there was a consensus that there is great need for more time and support for reflection.

Despite the sense of commitment to the program that the Coordinators perceived from school staff, there was also a sense that, even though they initially applied voluntarily for the program, at least some schools may require reaffirmation of the program’s effectiveness to keep them invested. This may be a reflection of frustrations growing from the degree of investment needed to meet the challenges of this complex initiative. This situation presents something of a “catch 22,” however: since initiative-wide impacts are often not observable in the early years, it is essential that schools remain invested long enough to achieve a strong level of implementation so that they have a chance to see the fruits of their efforts. Case studies that have begun as of the third year of the program evaluation could help provide greater insights into program impacts at the classroom and student level. If the findings are positive, this could help maintain participants’ investment in the program; although it will also be important for staff to view any findings of limitations as opportunities for improvement.

These findings, derived from conversations with program leadership as well as from the fidelity scales, would seem to call for program management to further explore the impediments that have been encountered and possible solutions that might be employed. District Coordinators cited a need for school administration to send a stronger message of buy-in to the program, making their expectations clearer and establishing accountability for implementing the program. Some suggested that, as part of this message, principals could spend more time in the classrooms.

This variability in implementation during Year 2 also has important implications for the current evaluation design. There are several reasons that might explain why a program impact was not detected in the current analysis. Obviously, the possibility that the program does not work as expected is one of them—otherwise there would be no need for a study. However, even if the program *were* effective, there are several reasons why an impact may have gone undetected. One is that the initiative is still in its early stages—especially among Cohort 2 schools—and has not yet had sufficient time to mature into a program that fully reflects the intended model. In some cases, for example, CPS’s area instructional managers may not be up to speed on the project and may perceive (rightly or wrongly) that other district mandates are not fully aligned with it, yet they are still obligated to enforce those mandates. Similarly, responsibilities of district staff (including the Program Director and Coordinators) also include some roles, such as capacity building, that are related to the Striving Readers Initiative, but compete for their time with more “direct” responsibilities such as providing training and conducting staff meetings. Similarly, some of the Coordinators noted that the principals are also struggling with balancing multiple initiatives, and suggested that they may need additional professional development. It is possible that with additional training, as all levels of staff increase their comfort level with the initiative, the connections between Striving Readers and other district initiatives will become more apparent, and will come to be seen more as complimentary than competing with Striving Readers. In the mean time, all of the variations and areas of limited implementation represent dilutions of the intended model that would be expected to reduce the likelihood of detecting a program effect experimentally through an intent-to-treat model.

A closely related reason for the difficulty in detecting program impact is that in an intent-to-treat model, results for all students and schools are considered on equal footing, including for schools that have so far been the least successful in implementing the model and, indeed, even among students who are not fully participating in the program. Results of the intent-to-treat model speak to whether the program is working as implemented—in this context at this time—but they do not necessarily reflect whether the intended *program model* would work if implemented correctly. Of course, to the extent that lack of findings in the intent-to-treat model is a reflection of imperfect implementation of the model, this again is a call to program management to continue to explore and address the factors that hinder implementation.

However, imperfect implementation is not the only reason that the evaluation might fail to detect program impacts even for an effective program model. One final factor that might reduce the chances of detecting impact is that, even in a rigorously controlled experimental design, findings can be confounded by threats to construct validity. First, a true experimental design—wherein the control group receives a placebo in place of the intervention—is probably impossible (and undesirable) to implement in human subjects research. Indeed, since some of the components of the Striving Readers model stem from previous initiatives at the district level (some of which were developed by the same university consultant), they are not entirely unique to the treatment

schools, and at least some of the district level training that is available to all schools is conducted by District Coordinators, and is therefore directly informed by the Striving Readers initiative. There is also the additional possibility that any differences that do exist between treatment and control groups are being further diminished by “compensatory rivalry” (Cook & Campbell, 1979): the control schools are quite aware that they are being compared to the Striving Readers schools and they may be making an extra effort toward their literacy curriculum because they do not want to be perceived as less successful. Even if their efforts are not motivated by a sense of competition, any educators worth their salt would seek out any promising avenue for improving instruction—including what they may learn from networking with staff from other schools, which is a common practice in Chicago. Indeed, at least one of the Striving Readers school coordinators has commented that she was aware of some of the control schools doing just that. Some control schools have, on their own, acquired some of the Striving Readers “targeted materials” such as listening centers and text sets, and have also adopted some of the specific instructional methods.

While there is obviously little that the district can do about the latter situation (indeed, the district would probably not *want* to discourage other schools from trying harder), it might still be possible (albeit more difficult) to detect a program effect, since an important difference that would still be expected to occur between treatment and control schools is in the way that resources are applied and in the quality with which Striving Readers schools utilize these methods. It is clear therefore that further efforts to strengthen program implementation could increase the likelihood of affecting more positive outcomes for students, while also improving the chances of detecting the initiative’s potential to impact student achievement. In addition, resources permitting, additional analyses that controlled for dosage and fidelity of implementation, even though they could not be conducted in an experimental design, could provide an alternate focus that would better isolate the impacts of the program in cases when it is implemented as designed.

## References

---

- Atwell, N. (1998). *In the middle: New understanding about writing, reading, and learning*. Boynton/Cook Publishers, Inc.
- Biancarosa, G., & Snow, C. E. (2004). *Reading Next — A vision for action and research in middle and high school literacy* (A Report to Carnegie Corporation of New York). Washington, DC: Alliance for Excellent Education.
- Banilower, E.R., Boyd, S.E., Pasley, J.D., Weiss, I.R. (2006). *Lessons from a decade of mathematics and science reform: A capstone report for the local systemic change through teacher enhancement initiative* [prepared for the National Science Foundation]. Chapel Hill, NC: Horizon Research, Inc.
- Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis for field settings*. Boston: Houghton Mifflin Company.
- Dede, C. (Ed.). (1998). *Learning with technology* (1998 ASCD Yearbook). Alexandria, VA: Association for Supervision and Curriculum Development.
- DeStefano, L, Hammer, V, Fiedler, E, & Downs, H (2006). *Evaluation of the implementation of the Illinois learning standards*. Urbana-Champaign, Illinois: University of Illinois College of Education.
- Illinois State Board of Education (ISBE) - Division of Assessment. (2007). *Illinois Standards Achievement Test: 2007 Technical Manual*. Retrieved January 15, 2009, Web site: [http://www.isbe.state.il.us/assessment/pdfs/isat\\_tech\\_2007.pdf](http://www.isbe.state.il.us/assessment/pdfs/isat_tech_2007.pdf)
- Illinois State Board of Education. (2008). *Illinois Standards Achievement Exam: Guide to the 2008 Illinois State Assessment Reading Mathematics Science*. USA: Pearson Education, Inc.
- Moats, L. (2001). When older kids can't read. *Education Leadership*.
- Moore, D.W., Alvermann, D.E., & Hinchman, K.A. (Eds.). (2000). *Struggling adolescent readers: A collection of teaching strategies*. International Reading Association.
- Peterson, C.L., Caverly, D.C., Nicholson, S.A., O'Neal, S., & Cusenbary, S. (2000). *Building reading proficiency at the secondary school level: A guide to resources*. San Marcos, TX: Southwest Texas State University and the Southwest Development Laboratory.
- Snow, C.E. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. RAND Reading Study Group. Santa Monica, CA: RAND Corporation.



## **APPENDICES**