

National Assessment of Title I: Interim Report

Executive Summary

National Assessment of Title I Interim Report

Executive Summary

Institute of Education Sciences

National Center for Education Evaluation and Regional Assistance
U.S. Department of Education

NCEE 2006-4000
February 2006

U. S. Department of Education

Margaret Spellings

Secretary

Institute of Education Sciences

Grover J. Whitehurst

Director

National Center for Education Evaluation and Regional Assistance

Phoebe Cottingham

Commissioner

February 2006

This report is in the public domain. Authorization to reproduce it in whole or in part is granted. While permission to reprint this publication is not necessary, the citation should be: *National Assessment of Title I Interim Report: Executive Summary*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, 2006.

To order copies of this report,

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 1398, Jessup, MD 20794-1398.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.
- Fax your request to 301-470-1244.
- Order online at www.edpubs.org.

This report also is available on the Department's Web site at <http://www.ed.gov/ies>.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

Preface

This is one of three documents constituting the mandated Interim Report on the National Assessment of Title I. This *Executive Summary* describes the studies that comprise the National Assessment of Title I and provides executive summaries of the findings of Volumes I and II. Volume I, *Implementation of Title I*, was prepared by the Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. Volume II, *Closing the Reading Gap: First Year Findings from a Randomized Trial of Four Reading Interventions for Striving Readers*, was prepared by the Corporation for the Advancement of Policy Evaluation.

Contents

I. Introduction	1
II. Executive Summary of Volume I: Implementation of Title I.....	9
III. Executive Summary of Volume II: Closing the Reading Gap	27
Appendix A. Independent Review Panel Members.....	46

Introduction

The Title I program began in 1965 as part of the Elementary and Secondary Education Act (ESEA) and is intended to help ensure that all children have the opportunity to obtain a high-quality education and reach proficiency on challenging state standards and assessments. As the largest federal program supporting elementary and secondary education (funded at \$12.7 billion in FY 2006), Title I, Part A targets these resources primarily to high-poverty districts and schools, where the needs are greatest. Title I provides flexible funding that may be used to provide additional instructional staff, professional development, extended-time programs, and other strategies for raising student achievement. The program focuses on promoting schoolwide reform in high-poverty schools and ensuring students' access to scientifically based instructional strategies and challenging academic content. Title I holds states, school districts, and schools accountable for improving the academic achievement of all students and turning around low-performing schools, while providing alternatives to students in such schools to enable them to receive a high-quality education.

The No Child Left Behind Act of 2001 (NCLB), which first went into effect beginning with the 2002-03 school year, reauthorized the Title I program and made a number of significant changes. NCLB strengthened the accountability provisions of the law, requiring that states establish assessments in each grade from 3-8 and once in grades 10-12, and set annual targets for school and district performance that would lead to all students reaching proficiency on those assessments by the 2013-14 school year. Schools and districts that do not make adequate yearly progress (AYP) towards this goal are identified as needing improvement and are subject to increasing levels of interventions designed to improve their performance and provide additional options to their students. NCLB also required that all teachers of core academic subjects become highly qualified, which the law defines as holding a bachelor's degree and full state certification, as well as demonstrating competency, as defined by the state, in each core academic subject that he or she teaches. These and other changes were intended to increase the quality and effectiveness not only of the Title I program, but of the entire elementary and secondary education system in raising the achievement of all students, particularly those with the lowest achievement levels.

A. National Assessment of Title I

As part of NCLB, the Congress mandated a National Assessment of Title I (Section 1501) to evaluate the implementation and impact of the program. This mandate also required the establishment of an Independent Review Panel (IRP) to advise the Secretary on methodological and other issues that arise in carrying out the National Assessment and the studies that contribute to this assessment. In addition, the law specifically requires a longitudinal study of Title I schools to examine the implementation and impact of the Title I program.

On November 6, 2002, the President signed the "Education Sciences Reform Act of 2002," establishing a new National Center for Education Evaluation and Regional Assistance (NCEE) in the Institute of Education Sciences. Part D of this Act assigned responsibility for the National Assessment of Title I to NCEE. The creation of this Center represented an important shift in the purposes of program evaluation and the types of methodology used in Department evaluation studies from broader policy and program assessments to specific scientific evaluations of program effectiveness.

In the past, Department program evaluation studies of Title I have, for the most part, focused on broader issues of program implementation, such as targeting of federal resources, compliance with federal laws and regulations, characteristics of program participants, and types of services provided. Such studies, now carried out by the Policy and Program Studies Service (PPSS) in the Office of Planning, Evaluation and Policy Development, include examinations of issues such as trends in student achievement, implementation of State assessment systems, accountability and support for school improvement, Title I school choice and supplemental educational services, teacher quality, and the targeting of federal Title I funds.

However, school superintendents, principals, and teachers often do not have the information they need in order to make sound decisions to improve instruction and raise student achievement. In many areas, the scientific evidence on the effectiveness of education programs is weak, inconsistent, or nonexistent. Evidence is needed on the effectiveness of specific interventions to inform Title I program improvement. NCLB repeatedly emphasizes the importance of adopting scientifically proven educational practices and programs. In an effort to significantly raise the quality of scientific evidence on program effectiveness, NCEE has launched a generation of evaluation studies that use the most rigorous evaluation designs possible to detect the impact of educational practices and programs on student achievement. Under the National Assessment of Title I, NCEE has begun studies of remedial reading programs, reading comprehension programs, and mathematics curricula to assess the effectiveness of educational programs in these important areas of academic achievement. These studies are using randomized trials in which schools or teachers are randomly assigned to an educational program or to the control condition. Such experimental designs are the most reliable and accurate way of estimating the effectiveness of an educational intervention.

This combination of implementation studies conducted by PPSS and effectiveness studies conducted by NCEE will provide valid evidence upon which to improve Title I services and the academic achievement of students. Implementation studies can provide nationally representative data on the types of programs and practices that schools have adopted. Effectiveness studies can provide evidence about which of those practices produce the best results. Together these two types of studies can provide the information needed to effectively target technical assistance and assist policymakers in making decisions on the best use of resources.

B. Independent Review Panel for the National Assessment of Title I

The mandated function of the Independent Review Panel (IRP) for the National Assessment of Title I is to advise on methodological and other issues that arise in carrying out the assessment. The IRP is to ensure that the assessment and studies adhere to the highest possible standards of quality with respect to research design, statistical analysis, and the dissemination of findings; and that the studies use valid and reliable measures to document program implementation and impacts. The IRP was appointed in November 2002 and is made up of researchers, education practitioners, parents, and members of other organizations involved with the implementation and operation of programs under Title I. A list of IRP members and their affiliations is included in Appendix A.

The IRP first met in January 2003 and has been instrumental in shaping the direction of implementation and effectiveness studies under the National Assessment of Title I. At this meeting, the IRP noted that an evaluation of the impact of Title I funds on student achievement was not feasible because it would require random assignment of Title I funds to eligible districts and schools. Past evaluations of activities supported by Title I have provided little information on how to improve student achievement. The IRP

recommended that Title I effectiveness studies focus on “what works” evaluations of well-defined interventions for improving achievement of high-poverty students in the critical areas of reading and mathematics. These evaluations would provide information on the effectiveness of specific interventions that could be adopted by schools to improve academic achievement. Additional information on IRP recommendations for effectiveness studies is included below in descriptions of each of these studies.

The IRP has also provided essential advice on the conduct of implementation studies. At its first meeting, the panel agreed that the mandated national longitudinal study of Title I schools should be launched as soon as possible, and most members advised that it should focus on program implementation rather than the impact of federal funds for reasons described above. However, the IRP also recommended that the study include an analysis, using a quasi-experimental design, of student achievement in schools that have been identified for improvement. The panel noted that although the longitudinal study (now known as the National Longitudinal Study of No Child Left Behind (NLS-NCLB)) is to focus on Title I schools, it should include a comparison group of non-Title I schools. The IRP recommended that the study include a survey of parents concerning Title I school choice and supplemental educational services provisions and other aspects of Title I, and provided advice regarding the study designs and data collection instruments for both the NLS-NCLB and a companion state-level study.

The IRP has met six times over the past three years. Several meetings were held in the first year after the panel’s appointment in November 2002: January 30-31, 2003; March 17-18, 2003; September 22, 2003; and November 9, 2003. There were also IRP meetings on November 21, 2004 and on July 29, 2005. The IRP has provided valuable advice on the design and implementation of the Title I studies as well as extensive comments on this *Executive Summary* and Volume I of the Interim Report.

C. Title I Implementation Studies

To answer questions of program implementation, the Department will rely on surveys of states, districts, schools, and teachers as well as more in-depth case studies and analyses of state performance reports and other extant data sources. Findings from these kinds of studies are valuable to the Congress, the Department, and educators as they assess the degree to which federal programs are being implemented as intended, describe the problems and challenges to implementation, and identify states and districts that have made significant progress.

The National Assessment’s two main data sources on NCLB implementation, the National Longitudinal Study of NCLB and the Study of State Implementation of Accountability and Teacher Quality under NCLB, both collected data in the 2004-05 school year, and preliminary findings from those studies are presented in this interim report. The report also includes data from earlier studies, state performance reports, and the National Assessment of Educational Progress (NAEP). These studies are listed below and a summary of their key findings on the implementation of Title I are reported later in this document. Volume I of this report contains an in-depth look at findings on the implementation of Title I.

1. National Longitudinal Study of NCLB (NLS-NCLB)

This study is examining the implementation of NCLB provisions concerning accountability, teacher quality, Title I school choice and supplemental services, and targeting and resource allocation. The study

is surveying districts, principals, classroom teachers, special education teachers, and Title I paraprofessionals in a nationally representative sample of 300 districts and 1,483 schools in the 2004-05 and 2006-07 school years. The study is also surveying parents and supplemental service providers in a small subsample of districts in both years. The study is collecting targeting and resource allocation data from all 300 districts in 2004-05 only. Finally, the study includes two exploratory achievement analyses that examine a) achievement outcomes for students participating in the Title I choice and supplemental services options in nine districts, and b) student achievement following identification of schools for improvement in two states.

2. Study of State Implementation of Accountability and Teacher Quality under NCLB (SSI-NCLB)

This companion study to the NLS-NCLB is collecting information from all states about their implementation of the accountability, assessment, and teacher quality provisions of the law, as well as Title III requirements for inclusion of students with limited English proficiency. The study is surveying state education staff responsible for implementing these provisions in 2004-05 and in 2006-07. In addition, the study is also analyzing extant data relating to state implementation, including state lists of schools and districts that did not make adequate yearly progress and those that were identified as in need of improvement.

3. Study of Title I Accountability Systems and School Improvement Efforts (TASSIE)

This study examines implementation of Title I accountability provisions during the transition years from 2001-02 (prior to implementation of NCLB) through 2003-04 (the second year of NCLB implementation). The study surveyed a nationally representative sample of 1,200 districts and 740 schools that had been identified for improvement under the previous authorization of ESEA.

4. Case Studies of the Early Implementation of Supplemental Educational Services

These case studies in nine districts examine the early experiences of districts implementing the NCLB supplemental services provisions in 2002-03 and 2003-04.

5. State Consolidated Performance Reports

These annual state reports, required under NCLB, provide data on student achievement on state assessments as well as basic descriptive information, such as numbers of identified schools and number of student participants.

6. National Assessment of Educational Progress (NAEP)

NAEP provides information on a common assessment for populations targeted by Title I.

The final report will provide more complete data on Title I implementation and outcomes, including information about the targeting and uses of Title I funds, services for private school students, findings from the NLS-NCLB parent survey and supplemental service provider survey. Quasi-experimental analyses of student achievement related to participation in the Title I choice and supplemental services options, as well as the identification of schools for improvement, will be included in the final report.

D. Title I Effectiveness Studies

A central principle of NCLB is that states, districts, schools, and teachers adopt instructional practices backed by evidence of effectiveness from scientifically based research. This principle has created a demand for rigorous evaluation evidence currently unavailable for most education programs and instructional areas. For this reason, the Department's evaluation strategy for Title I features a strong emphasis on evaluation studies that are designed to produce rigorous scientific evidence on the effectiveness of specific education programs and practices that are critical to the effective use of Title I funds.

At the second meeting of the Independent Review Panel on March 17-18, 2003, presentations were made by reading and mathematics experts on what we know and need to know in these areas. Ultimately, three large-scale evaluations were undertaken. The first is examining the effects of remedial reading programs for 3rd and 5th graders. Based on the advice of an expert panel formed by NCEE, the second evaluation will look at the effectiveness of reading comprehension interventions for 5th graders. The third evaluation will assess the effectiveness of mathematics curricula that are widely used in the early elementary grades. The rationales for these three large-scale evaluations of specific interventions are described briefly below.

1. Remedial Reading Interventions

According to the NAEP,¹ nearly 4 in 10 fourth graders read below the basic level. Historically, nearly three-quarters of these students never attain average levels of reading skill. To address this problem, many school districts have created remedial programs that aim to improve the skills of students reading below grade level. However, it is very difficult for these students to close the reading gap and become average readers. We know very little about the effectiveness of remedial reading programs for struggling readers in regular school settings.

Closing the Reading Gap, the evaluation of remedial reading programs, is addressing three broad questions:

- *What is the impact of being in any of four promising remedial reading interventions, considered as a group, relative to the instruction provided by the schools? What is the impact of being in one of the remedial reading programs that focuses primarily on developing word-level skills, considered as a group, relative to the instruction provided by the schools? What is the impact of being in each of the four particular remedial reading interventions, considered individually, relative to the instruction provided by the schools?*
- *Do the impacts of programs vary across students with different baseline characteristics?*
- *To what extent can the instruction provided in this study close the reading gap and bring struggling readers within the normal range, relative to the instruction provided by their schools?*

The key findings from the first report on *Closing the Reading Gap* are summarized later in this document. Volume II of this Interim Report contains the full report on this study. Future reports will include the longer term impacts of these interventions on student achievement.

2. Reading Comprehension Interventions

The decision to conduct an evaluation of the efficacy of reading comprehension interventions for informational materials in content areas such as social studies or science resulted from a series of

discussions between the IRP and reading experts, as well as from the advice of a subsequent expert panel convened to identify important and policy-relevant evaluation questions to study in reading. The expert panel's advice was that there are increasing cognitive demands on student knowledge in middle elementary grades where students become primarily engaged in reading to learn, rather than learning to read. Children from disadvantaged backgrounds lack general vocabulary as well as vocabulary related to academic concepts that enable them to comprehend what they are reading and acquire content knowledge. They also do not know how to use strategies to organize and acquire knowledge from informational text in content areas such as science and social studies.² The panel advised that strategies for improving comprehension are not as well developed as those for decoding and fluency. While there are multiple techniques for direct instruction of comprehension in narrative text that have been well-demonstrated in small studies, there is not as much evidence on teaching reading comprehension within content areas.

This evaluation of reading comprehension is addressing the following questions:

- *Can promising reading comprehension interventions improve student reading achievement of informational text?*
- *What are the most effective reading comprehension interventions for improving student reading achievement of informational text?*
- *Under what conditions and practices do reading comprehension interventions improve student reading achievement of informational text?*

Five supplemental interventions have been selected by an expert panel and are being piloted in 5th grade classrooms during the 2005-06 school year. Selection of the interventions was based on existing research evidence, quality of the intervention design, capability to implement the intervention, and appropriateness of the intervention for the target population. All of the selected interventions teach reading comprehension of text containing information such as science or social studies content. The interventions being piloted and their publishers are:

- **CRISS** (Project CRISS): CRISS teaches a wide array of comprehension and note-taking strategies using science text. Students then apply the strategies to the actual texts used in their social studies and science classes. The program teaches students the difference between reading a text for basic information, reading for understanding a physical or natural phenomenon, and how to create succinct summaries. It also stresses active reading strategies such as asking oneself questions while reading and then jotting down the answers. The program is designed to be used for 30 minutes each day during language arts, science, or social studies periods. Teachers participate in three days of initial training and one day of follow-up training.
- **ReadAbout** (Scholastic): Students are taught reading comprehension skills such as author's purpose, main idea, cause and effect, compare and contrast, summarizing, and inferences primarily through a computer program. The program is designed to be used for 30 minutes per day, and students apply what they have learned during this time to a selection of science and social studies trade books. Teachers receive two days of initial training plus two additional days during the school year.
- **Read for Real** (Chapman University; Zaner-Bloser): In Read for Real, teachers use a six-volume set of books to teach reading strategies appropriate for before, during and

after reading such as previewing, activating prior knowledge, setting a purpose, main idea, graphic organizers, and text structures. Students use the materials for 30 to 45 minutes per day. Each of these units includes vocabulary, fluency, and writing activities. Teachers participate in three days of initial training and two, one-day follow-up training sessions.

- **Reading for Knowledge** (Success for All Foundation): Reading for Knowledge, a 30-minute daily program, makes extensive use of cooperative learning strategies and a process called SQRRL (Survey, Question, Read, Restate, Review, Learn). Teachers receive 2.5 days of initial training in addition to monthly follow up sessions.
- **Skills Handbooks** (SRA): Students are taught reading comprehension skills (e.g., compare and contrast, cause and effect, fact and opinion, main idea, summarizing) through a set of workbooks and practice activities in science and social studies. Teachers are expected to use the program for about 30 minutes per day. They receive one day of initial training and an additional 16 hours during the school year.

The reading comprehension interventions that are successfully piloted will be randomly assigned to a total of 100 participating elementary schools during the 2006-07 school year. The impact of the interventions on reading comprehension of informational social studies texts will be assessed. The first report on the effectiveness of the reading comprehension interventions is planned for Spring 2008.

3. Mathematics Curricula

The decision to conduct an evaluation of the effectiveness of mathematics curricula resulted from a series of discussions with and recommendations from the IRP, the Office of Elementary and Secondary Education (OESE), and an expert panel convened to provide advice on policy-relevant questions it would be important to address in an impact evaluation focused on mathematics. Information on the effectiveness of mathematics curricula is crucial to improving performance on state mathematics assessments under NCLB. There is considerable controversy about what mathematics children should learn and how it should be taught, but there is very little reliable information available to educators and policy makers about which curricula are most likely to improve mathematics achievement.³

This evaluation will focus on early elementary grades because disadvantaged children fall behind their more advantaged peers in basic mathematics competencies even before entering elementary school. If basic concepts are not mastered in early elementary grades, students have great difficulty understanding more advanced mathematics concepts in upper elementary grades. The evaluation will compare different approaches to teaching early elementary mathematics, since there are many mathematics curricula that are being widely implemented without evidence of their effectiveness.

- *What is the relative effectiveness of a variety of mathematics curricula on mathematics achievement for early elementary school students in disadvantaged schools?*
- *Under what conditions is each mathematics curriculum most effective?*

Up to five mathematics curricula will be competitively selected during the 2005-06 school year. The selected curricula will be randomly assigned to participating schools. Teachers will be trained and the curricula will be implemented during the 2006-07 school year. Data will be collected on implementation of each curriculum and student mathematics achievement at the beginning and end

of the 2006-07 school year. The first report on the relative effectiveness of the mathematics curricula is planned for Spring 2008.

Organization of This *Executive Summary*

The following two sections of this document contain the executive summaries from Volume I and Volume II of the *National Assessment of Title I: Interim Report*. Volume I contains key findings on the implementation of Title I under No Child Left Behind. Volume II is a report on the findings from *Closing the Reading Gap*, an evaluation of the impact of supplemental remedial reading programs on student achievement.

End Notes

¹ *The Nation's Report Card: Reading 2005 (2005)*. Washington, DC: US Department of Education.

² National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel, Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction* (NIH Publication No. 00-4769). Washington, DC: US Government Printing Office.

³ The National Academy of Sciences (2004). *On Evaluating Curricular Effectiveness: Judging the Quality of K-12 Mathematics Evaluations*. Washington, DC: The National Academies Press.

Executive Summary of Volume I: Implementation of Title I

A Report Prepared for the Institute of Education Sciences
by Stephanie Stulich, Elizabeth Eisner, Joseph McCrary, and Collette Roney
Policy and Program Studies Service
Office of Planning, Evaluation and Policy Development

The Title I program began in 1965 as part of the Elementary and Secondary Education Act (ESEA) and is intended to help ensure that all children have the opportunity to obtain a high-quality education and reach proficiency on challenging state standards and assessments. The No Child Left Behind Act of 2001 (NCLB) built upon and expanded the assessment and accountability provisions that had been enacted as part of the ESEA's previous reauthorizing legislation, the Improving America's Schools Act (IASA), while also creating new provisions related to parental choice and teacher quality. These and other changes were intended to increase the quality and effectiveness not only of the Title I program, but also of the entire elementary and secondary education system in raising the achievement of all students, particularly those with the lowest achievement levels.

As part of the No Child Left Behind Act, the Congress mandated a National Assessment of Title I to evaluate the implementation and impact of the program. The mandate specifically requires a longitudinal study of Title I schools, as well as an Independent Review Panel composed of expert researchers and practitioners to advise the U.S. Department of Education on the conduct of the National Assessment. An interim report is due in 2005 and a final report is due in 2007.

This report constitutes Volume I of the National Assessment of Title I interim report and focuses on implementation of key Title I provisions related to state assessments, accountability, school choice and supplemental educational services, and teacher quality, as well as examining trends in student achievement. The report draws on data from two evaluations of NCLB implementation conducted by the Department, the National Longitudinal Study of NCLB and the Study of State Implementation of Accountability and Teacher Quality Under NCLB, both of which collected data in the 2004-05 school year. The report also includes data from earlier studies, state performance reports, the National Assessment of Educational Progress, and other sources.

The final report will provide more complete data on Title I implementation and outcomes, including information about the targeting and uses of Title I funds, services for private school students, findings from a parent survey about parents' experiences with choice options, and analyses of a) student outcomes associated with participation in the Title I choice and supplemental services options and b) the impact on student achievement of identifying schools for improvement.

A. Key Provisions of Title I under the No Child Left Behind Act

NCLB, which went into effect beginning with the 2002-03 school year, strengthened the assessment and accountability provisions of the law, requiring that states annually test all students in grades 3-8 and once in grades 10-12 on assessments that are aligned with challenging state standards. States must also set targets for school and district performance that lead to all students achieving proficiency on state reading and mathematics assessments by the 2013-14 school year. Schools and districts that do not make adequate yearly progress (AYP) towards this goal for two consecutive years are identified as needing improvement and are subject to increasing levels of interventions designed to improve their performance, as well as provide additional options to their students. In schools identified for improvement, districts must offer students the option to transfer to another school. If an identified school misses AYP again (for a third year), low-income students in the school must be offered the option to receive

supplemental educational services from a state-approved provider. If an identified school misses AYP for a fourth year, the district must take one of a set of “corrective actions” specified in the law, and if the school misses AYP for a fifth year, the district must begin planning to restructure the school.

NCLB also requires that all teachers of core academic subjects become “highly qualified,” which the law defines as having a bachelor’s degree and full state certification as well as demonstrating competency, as defined by the state, in each core academic subject that they teach. Exhibit E-1 provides a more detailed summary of key NCLB provisions.

Exhibit E-1
Key Provisions of the No Child Left Behind Act

State assessments	States must implement annual state assessments in reading and mathematics in grades 3-8 and at least once in grades 10-12, and in science at least once in each of three grade spans: 3-5, 6-9, and 10-12. Assessments must be aligned with challenging state content and academic achievement standards. States must provide for participation of all students, including students with disabilities and limited English proficient (LEP) students. States must provide for the assessment of English language proficiency of all LEP students.
Adequate yearly progress (AYP)	States must set annual targets that will lead to the goal of all students' reaching proficiency in reading and mathematics by 2013-14. For each measure of school performance, states must include absolute targets that must be met by key subgroups of students (major racial/ethnic groups, low-income students, students with disabilities, and LEP students). Schools and districts must meet annual targets for each student subgroup in the school, and must test 95% of students in each subgroup, in order to make "adequate yearly progress." States also must define an "other academic indicator" that schools must meet in addition to proficiency targets on state assessments.
Schools identified for improvement	Schools and districts that do not make AYP for two consecutive years are identified for improvement and are to receive technical assistance to help them improve. Those that miss AYP for additional years are identified for successive stages of interventions, including corrective action and restructuring (see below). To leave "identified for improvement" status, a school or district must make AYP for two consecutive years.
Public school choice	Districts must offer all students in identified schools the option to transfer to a non-identified school, with transportation provided by the district.
Supplemental educational services	In schools that miss AYP for a third year, districts also must offer low-income students the option of supplemental educational services from a state-approved provider.
Corrective actions	In schools that miss AYP for a fourth year, districts also must implement at least one of the following corrective actions: replace school staff members who are relevant to the failure to make AYP; implement a new curriculum; decrease management authority at the school level; appoint an outside expert to advise the school; extend the school day or year; or restructure the internal organization of the school.
Restructuring	In schools that miss AYP for a fifth year, districts also must begin planning to implement at least one of the following restructuring interventions: reopen the school as a charter school; replace all or most of the school staff; contract with a private entity to manage the school; turn over operation of the school to the state; or adopt some other major restructuring of the school's governance. Districts must spend a year planning for restructuring and implement the school restructuring plan the following year.
Highly qualified teachers	All teachers of core academic subjects must be "highly qualified" as defined by NCLB and the state. To be highly qualified, teachers must have a bachelor's degree, full state certification, and demonstrated competence in each core academic subject that they teach. Subject-matter competency may be demonstrated by passing a rigorous state test, completing a college major or coursework equivalent, or (for veteran teachers) meeting standards established by the state under a "high, objective uniform state standard of evaluation" (HOUSSE).

B. Profile of Title I Participants and Resources

Funding for Title I, Part A, has increased by 46 percent over the past five years, after adjusting for inflation, from \$7.9 billion in FY 2000 to \$12.7 billion in FY 2006.¹ Title I funds go to nearly all of the nation's school districts and to 55 percent of all public schools, but are more strongly targeted to high-poverty districts and schools than are state and local education funds.² Most Title I funds go to elementary schools, and three-fourths of Title I participants are in pre-kindergarten through grade 6.³

Fueled by a growing use of Title I schoolwide programs, the number of students counted as Title I participants has more than doubled in recent years, rising from 6.7 million in 1994-95 to 16.5 million in 2002-03 (a 146 percent increase). The dramatic increase in participation is due in part to the way that students are counted: when a school converts from targeted assistance to a schoolwide program, all students in the school are counted as Title I participants instead of just the lowest-achieving students who are receiving specific targeted services. In 2002-03, 84 percent of Title I participants were in schoolwide programs.⁴

C. Trends in Student Achievement

This report examines trends in student achievement using both state assessment data and the National Assessment of Educational Progress (NAEP). We also examine recent trends in graduation rates, another important indicator of student achievement.

Student achievement on state assessments represents the primary criterion that the Title I legislation applies to measure school success, but these data cannot be aggregated across states to examine national trends, because they measure different content and use different achievement levels. In addition, many states have revised their assessment systems in recent years, so they often do not have the trend data needed to assess student progress. This interim report examines recent three-year trends (2000-01 through 2002-03) in 23 states that had consistent assessments in place over this period; however, few states had these data available for all student subgroups during this period.

The NAEP provides a high-quality assessment that is consistent across states, making the data useful for examining national trends in student achievement. However, the NAEP is not aligned with individual state content and achievement standards, so it does not necessarily measure what students are expected to learn in their states. This report examines achievement trends on both the Main NAEP (1990 to 2005) and the Trend NAEP (1971 to 2004), with a focus on recent trends. The Main NAEP was created in the early 1990s to provide an assessment that is more consistent with current content focuses and testing approaches, while the Trend NAEP continues the original NAEP assessment begun in the 1970s in order to track long-term trends. In general, the Main NAEP places greater emphasis on open-ended and extended response items and less emphasis on multiple choice questions. In addition, the Main NAEP reports on the percentages of students performing at various achievement levels (Basic, Proficient, and Advanced) as well as average scale scores, while the Trend NAEP reports only scale scores. The National Center for Education Statistics (NCES) has stated that although results from these two NAEP assessments cannot be compared directly, comparisons of the patterns they show over time, especially for student demographic groups, may be informative.

For both state assessment and NAEP results, recent achievement trends through 2004 or 2005 are positive overall and for key subgroups. At this early stage of NCLB implementation— states, districts, and schools only began to implement the NCLB provisions in 2002-03—it is too early to say whether these trends are attributable to NCLB, to other improvement initiatives that preceded it, or a combination of both. The data presented below provide a baseline indicator of achievement levels and trends that existed at the time that NCLB implementation began. They may very well reflect pre-existing state standards-based reform efforts and accountability systems that NCLB was intended to strengthen. Moreover, even when additional years of assessment data become available, such data will be limited in their ability to precisely address the impact of NCLB, because it is difficult to separate the impact of NCLB from the effects of other state and local improvement efforts.

1. Student Achievement on State Assessments

Are students whom Title I is intended to benefit (including low-income students, racial/ethnic minorities, LEP students, migrant students, and students with disabilities) making progress toward meeting state academic achievement standards in reading and mathematics?

In states that had three-year trend data available from 2000-01 to 2002-03, the percentage of students achieving at or above the state’s proficient level rose for most student subgroups in a majority of the states (see Exhibit E-2), but the increases in student proficiency were often small. For example, state reading assessments administered in the 4th grade or an adjacent elementary grade show achievement gains in elementary reading for low-income students in 12 out of 16 states. Across all student subgroups examined, states showed achievement gains in about three-fourths of the cases. Results for mathematics and for 8th grade show similar patterns.

Exhibit E-2		
Number of States Showing an Increase in the Percentage of 4 th -Grade Students Performing at or Above the State’s Proficient Level from 2000-01 to 2002-03, by Student Subgroup		
	Reading	Mathematics
All students	11 out of 23 states	17 out of 23 states
Low-income	12 out of 16 states	10 out of 10 states
Black	5 out of 7 states	5 out of 7 states
Hispanic	6 out of 7 states	5 out of 7 states
White	7 out of 7 states	7 out of 7 states
LEP	12 out of 20 states	15 out of 20 states
Migrant	11 out of 15 states	12 out of 16 states
Students with disabilities	14 out of 20 states	16 out of 20 states

Exhibit reads: The proportion of students performing at or above states’ “proficient” levels in 4th-grade reading (or another nearby elementary grade) increased from 2000-01 to 2002-03 in 11 out of 23 states that had consistent trend data available.

Note: For states that did not consistently assess students in 4th-grade reading and mathematics from 2000-01 to 2002-03, this table is based on either 3rd-grade or 5th-grade results.

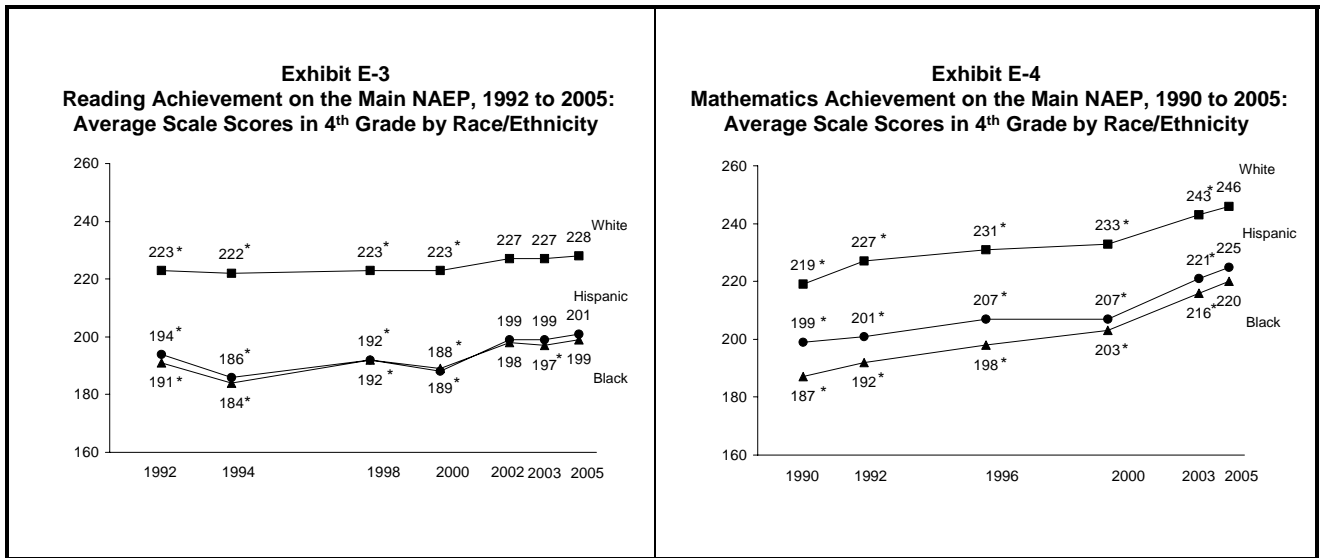
Source: Consolidated State Performance Reports (n = 23 states).

Based on trend data for 20 states, most would not meet the goal of 100 percent proficiency by 2013-14 unless the percentage of students achieving at the proficient level increased at a faster rate. For example, four out of 11 states with consistent elementary reading assessment data for low-income students would meet the 100 percent goal by 2013-14 for this subgroup if they sustained the same rate of growth that they achieved from 2000-01 to 2002-03. Looking across six different student subgroups (low-income, black, Hispanic, LEP, migrant, and students with disabilities), an average of 33 percent of the subgroups within these states would be predicted to reach 100 percent proficiency based on current growth rates.

2. Student Achievement on the National Assessment of Educational Progress (NAEP)

Are students, especially disadvantaged students, showing achievement gains on the National Assessment of Educational Progress?

Recent NAEP trends show gains in 4th-grade reading and especially in mathematics for black and Hispanic students and for students in high-poverty schools. For example, from 2000 to 2005, black students gained 10 points in reading and Hispanic students gained 13 points, while in mathematics, black students gained 17 points and Hispanic students gained 18 points. Over the longer term, black and Hispanic students showed even larger gains in mathematics (33 points and 26 points, respectively, from 1990 to 2005), but somewhat smaller gains in reading (eight points and seven points, respectively, from 1992 to 2005) (see Exhibits E-3 and E-4).



* Indicates that the score is significantly different from the one in 2005 (p<.05).

Source: National Center for Education Statistics, Main NAEP.

Eighth-grade students also made significant gains in mathematics but not in reading. The 12th-grade NAEP assessment was not administered in 2003 or 2005.

The long-term achievement trends measured by the Trend NAEP show significant gains for all three age groups tested in mathematics and for 9-year-olds and 13-year-olds in reading. In addition, recent gains from 1999 to 2004 are significant for 9-year-olds in both mathematics and reading and for 13-year-olds in mathematics. Black and Hispanic students show substantial gains on the Trend NAEP, both in the most recent period as well as over the full three decades covered by the assessment.

Are achievement gaps between disadvantaged students and other students closing over time?

State assessments and NAEP both provide some indications that achievement gaps between disadvantaged students and other students may be narrowing, but recent changes are small. For example, state assessments show a slight reduction in the achievement gap between low-income students and all students in most states, typically a reduction of one to three percentage points. On the Trend NAEP, achievement gains for black and Hispanic students since the 1970s substantially outpaced gains made by white students, resulting in significant declines in black-white and Hispanic-white achievement gaps, but recent changes in achievement gaps often were not statistically significant.

3. Graduation Rates

Are graduation rates improving over time?

Under NCLB, high schools are held accountable for graduation rates, but methods for calculating graduation rates vary considerably across states. The averaged freshman graduation rate (calculated by NCES based on data from the Common Core of Data) is useful for providing a common standard against which state-reported graduation rates may be compared. The median state graduation rate in 2002 was 84 percent based on state reports and 75 percent based on the averaged freshman graduation rate.⁵

The recent trend in the averaged freshman graduation rate has been fairly level, and the mean graduation rate in 2002 (73 percent) was the same as in 1996.

D. Implementation of State Assessment Systems

1. Development of Assessments Required under No Child Left Behind

To what extent have states implemented the annual assessments in reading, mathematics, and science that will be required under NCLB?

While some states have standards and assessments in place in all of the required grade levels, most states need to implement additional assessments to meet the NCLB requirements by 2005-06 for reading and mathematics and by 2007-08 for science. As of March 2005, 27 states had completed their first full administration of all required reading assessments; 26 states had done so for all required mathematics assessments; and 22 states had done so for all required science assessments. Most of the remaining states had at least field-tested all of the required assessments.⁶

How are states developing their English language proficiency assessments?

Many state approaches to assessing English language proficiency (ELP) were still evolving as of 2004-05. All states had an assessment in place for 2004-05, but 44 states indicated that they anticipated making revisions to their ELP assessments. Twenty states reported that they had an ELP assessment in place that met NCLB requirements, 27 states plan to have an ELP assessment that meets NCLB requirements in place for 2005-06, and five states had not made a decision as to which ELP assessment instrument they will use in 2004-05.⁷

2. Inclusion and Accommodations

To what extent do state assessment systems include students with special needs?

Most states have met the requirement to annually assess 95 percent or more of their students, including major racial/ethnic groups, students with disabilities, limited English proficient (LEP) students, and low-income students. However, 14 states did not meet the minimum test participation requirement for one or more student subgroups. Ten states assessed fewer than 95 percent of one or more minority student groups (black, Hispanic, and/or Native American), and nine states did not meet the test participation requirement for LEP students.⁸

The lowest participation rates were for students with disabilities. While states missing the test participation requirement for other subgroups often missed by just one or two percentage points, states that failed to assess 95 percent of students with disabilities typically had lower participation rates for those students (as low as 77 percent in one state).

3. Disaggregated Student Achievement Data

How fully are states meeting NCLB requirements for reporting state assessment data?

The number of states that report student achievement data has more than doubled since NCLB was enacted. Fifty states present data disaggregated by race/ethnicity and gender and for limited English proficient students, students with disabilities, and low-income students on state report cards.⁹

E. Accountability and Support for School Improvement

1. School Identification for Improvement

What types of schools are identified for improvement?

States identified 13 percent of all schools for improvement for 2004-05. Of these, 9,028 were Title I schools (18 percent of Title I schools), representing nearly a 50 percent increase over the approximately 6,000 Title I schools identified for the previous two years (see Exhibit E-5). Most (76 percent) of the identified Title I schools were in their first year or second year of improvement, 12 percent were in corrective action, and 12 percent were in restructuring status. The number and percentage of Title I schools identified for improvement varied considerably across states.¹⁰

Schools in large and urban districts, and those with high concentrations of poor, minority, and LEP students, were more likely to be identified than other schools. For example, just over one-third of all schools with 75 percent or more of their students from low-income families or minority groups were identified schools in 2004-05, compared with fewer than 5 percent of schools with low concentrations of these students. Middle schools also were more likely to be identified (18 percent of middle schools) than were elementary or high schools (11 percent at each level). Ten percent of districts (or 1,511 districts) also were identified for 2004-05; 32 percent of these had no identified schools.¹¹

**Exhibit E-5
Number and Percentage of Identified Title I Schools,
1996-97 to 2004-05**

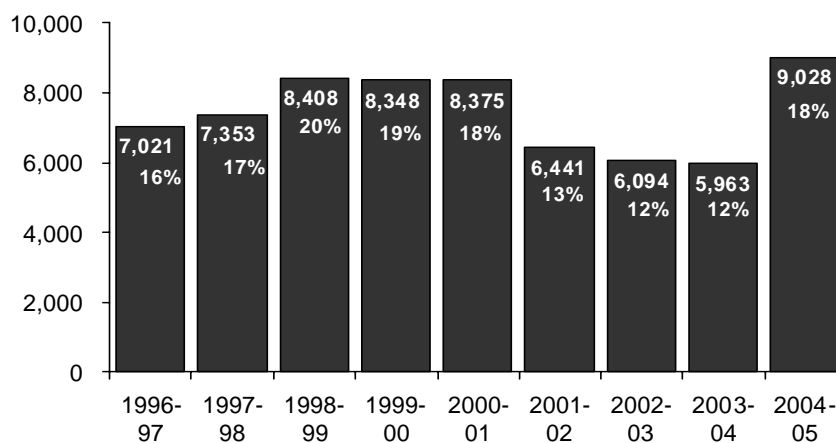


Exhibit reads: In 2004-05, 9,028 Title I schools had been identified for improvement based on test scores for 2003-04 and earlier years; these identified schools represented 18 percent of all Title I schools in that year.

Note: The first year that schools were identified for improvement based in part on NCLB AYP definitions was 2003-04, based on assessments administered in 2002-03. However, schools are identified when they miss AYP for two consecutive years, and 2004-05 was the first year that includes schools identified because they missed NCLB AYP targets for two consecutive years.

Sources: Consolidated State Performance Reports (1996-97 to 2002-03); Study of State Implementation of Accountability and Teacher Quality Under NCLB (2003-04 and 2004-05) (based on data from 50 states and the District of Columbia).

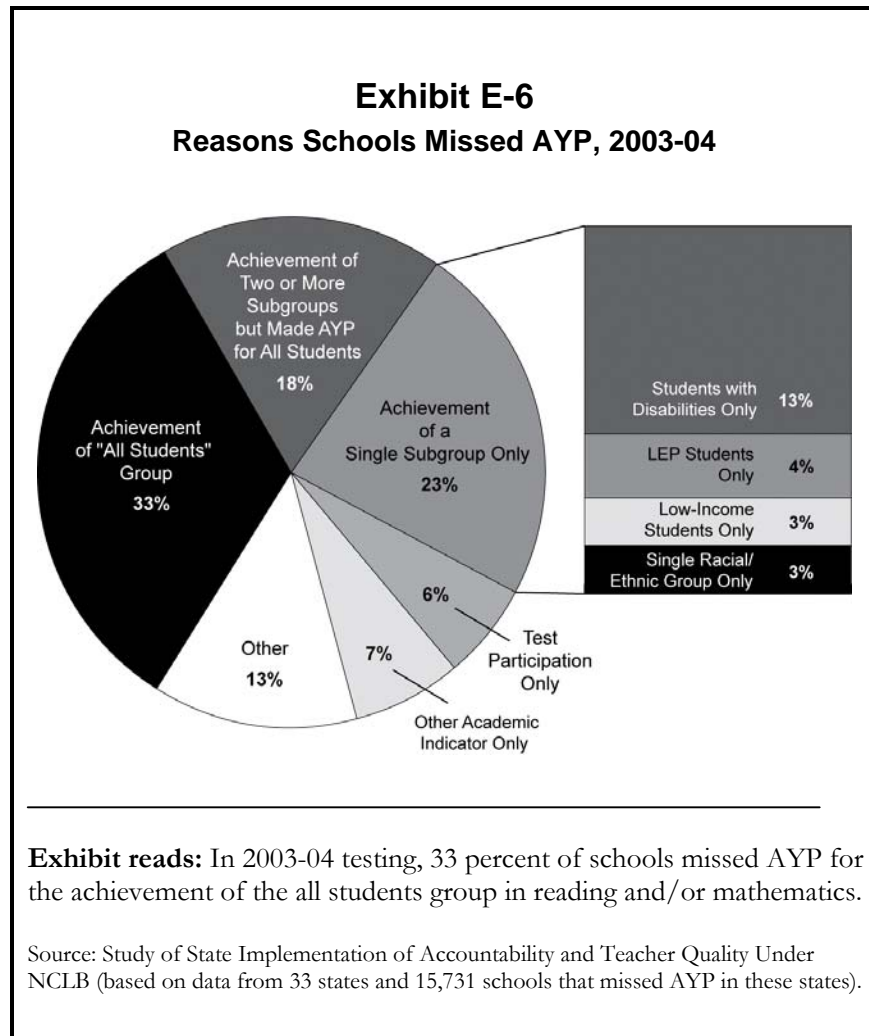
2. Adequate Yearly Progress

What are the reasons schools did not make adequate yearly progress (AYP)?

Three-fourths (75 percent) of all schools and 71 percent of districts met all applicable AYP targets in 2003-04 testing. The number of all schools missing AYP (21,540) based on 2003-04 testing is nearly double the number of schools identified for improvement for 2004-05 (11,530).¹² If many non-identified schools that did not make AYP in 2003-04 testing miss AYP again the following year, the number of identified schools could rise substantially in 2005-06.

Schools most commonly missed AYP for the achievement of all students and/or multiple subgroups; only in a minority of cases did schools miss only one AYP target. Based on data from 33 states, among schools that missed AYP in 2003-04, 33 percent did not meet achievement targets for the “all students” group in reading or mathematics, and another 18 percent missed AYP for the achievement of two or more subgroups (see Exhibit E-6). Only 23 percent missed AYP solely due to the achievement of a single subgroup.

Twenty percent missed AYP due to the “other academic indicator,” but only 7 percent missed for this indicator alone. More than one-fourth (29 percent) missed AYP due to insufficient test participation rates, but only 6 percent missed solely due to test participation. The remaining 13 percent of schools that missed AYP missed for other combinations of AYP targets.¹³



However, schools that were held accountable for more subgroups were less likely to make AYP. Among schools for which AYP was calculated for six or more subgroups, 39 percent did not make AYP, compared with 10 percent of schools for which AYP was calculated based on only one subgroup. More than one-fifth of those schools that were held accountable for the achievement of African-American students, LEP students, or students with disabilities did not make AYP for those subgroups in 2003-04 testing. Schools with subgroups of students from low-income families, Hispanic students, or Native American students were somewhat less likely to miss AYP for those subgroups (12 to 15 percent). Schools were much less likely to miss AYP due to the achievement of white or Asian students (1 percent and 4 percent of schools with these subgroups, respectively).¹⁴

3. School Improvement Activities

What assistance is provided to districts and schools identified for improvement? What interventions are implemented in these districts and schools?

All states notified schools about their identification status for 2004-05 based on 2003-04 testing, and a majority provided preliminary results before September 2004, but 20 states did not, and only 15 states provided final results by that time.¹⁵ NCLB regulations require states to notify schools and districts of their school improvement status prior to the beginning of the school year; this is important to enable districts with identified schools to notify parents of eligible students about their Title I choice options in a timely manner.

Identified schools were much more likely to report needing assistance in a variety of specific areas than non-identified schools, and they also reported receiving more days of assistance than non-identified schools. Identified schools were most likely to report needing assistance to improve the quality of teachers' professional development (80 percent), and most schools needing this assistance reported that they received it (91 percent). The most common improvement strategies implemented by identified schools included developing a school improvement plan, using assessment data to inform instruction, and providing additional instruction to low-achieving students.¹⁶

Nearly one-third (30 percent) of identified elementary schools reported increasing the amount of instructional time in reading by more than 30 minutes in 2004-05, and 17 percent reported a similar increase in instructional time for mathematics. Non-identified schools less frequently reported such increases. At the secondary school level, identified schools also more commonly reported increasing instructional time for low-achieving students in reading (55 percent).¹⁷

Almost all states had implemented a statewide system of support for identified schools by fall 2004, and these often involved school support teams and specialized individuals. Twenty-one states noted that an important objective of their statewide systems of support was to build district capacity to provide support to identified schools. Most states applied NCLB consequences for school identification (i.e., public school choice, supplemental services, corrective actions, and restructuring) to Title I identified schools only.¹⁸ Most states (42) reported that providing assistance to all schools identified for improvement was a moderate or serious challenge in 2003-04.¹⁹

Large and urban districts more commonly provided assistance of various kinds to identified schools than smaller districts. For example, in 2002-03, two-thirds of very large districts reported employing more than one full-time equivalent (FTE) staff member per identified school to provide assistance to those schools, compared with one-third of small districts.²⁰

Title I schools in corrective action status nearly universally experienced the interventions NCLB defines for schools in this stage of improvement. Corrective actions were implemented in 95 percent of Title I schools in corrective action status in 2004-05. The most common corrective actions experienced by Title I schools in this status in 2003-04 and 2004-05 resembled forms of technical assistance rather than sanctions. For instance, 90 percent of Title I

schools in corrective action were required to implement new research-based curricula or instructional programs and 58 percent had an outside expert appointed to advise the school.²¹

F. School Choice and Supplemental Educational Services

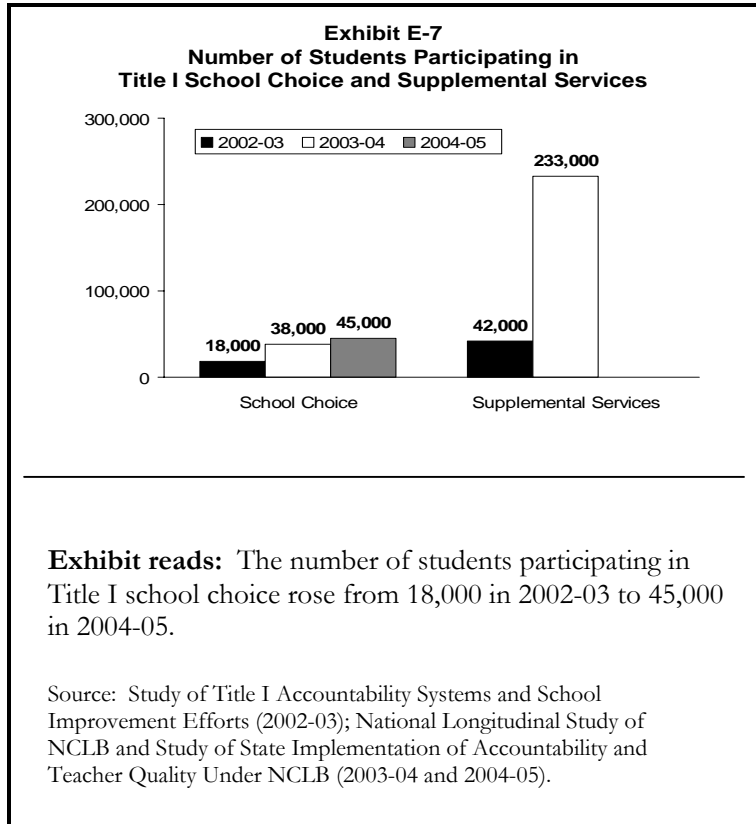
1. Eligibility and Participation

How many students are eligible to participate, and how many actually do so?

Although more students were eligible to participate in the Title I school choice option, a larger number actually participated in the supplemental services option.

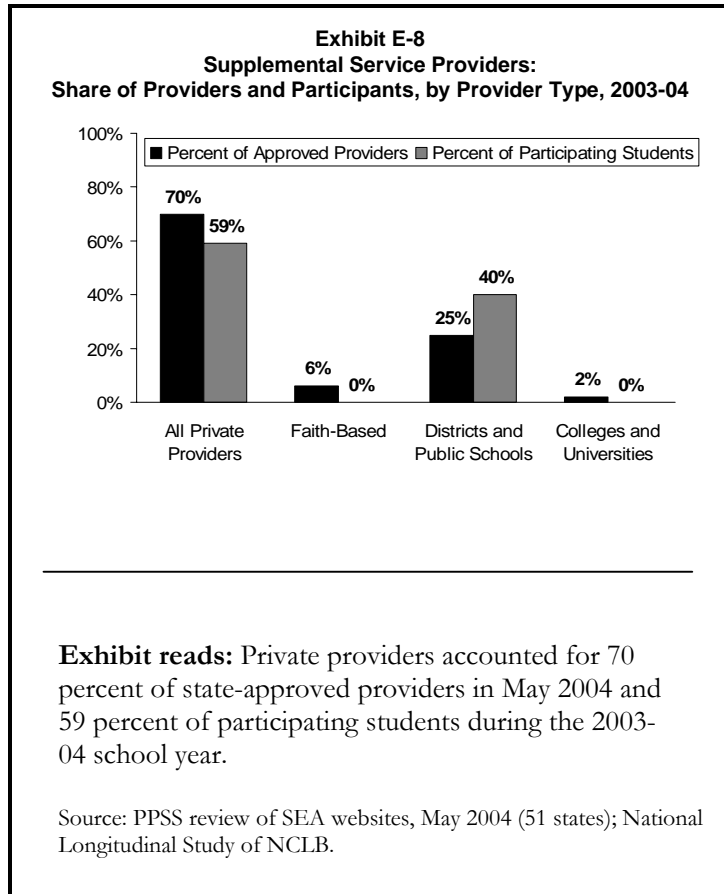
Based on district reports, twice as many students were eligible to transfer to another school under the Title I school choice option in 2003-04 (3.9 million) as were eligible to receive supplemental services (1.4 million). However, six times as many students actually participated in the supplemental services option (233,000) as participated in the school choice option (38,000) in that year (see Exhibit E-7).

The number of schools where supplemental services were offered tripled from 2002-03 to 2003-04 (from 800 to 2,500), while the number where Title I school choice was offered increased from 5,100 in 2002-03 to 6,200 in 2004-05. Title I school choice was offered in about 6,200 schools and 1,800 districts in 2004-05, and supplemental services were offered in 2,500 schools and 500 districts in 2003-04.²²



The number of state-approved supplemental service providers has tripled over the past two years, rising from 997 in May 2003 to 2,734 in May 2005.

Private firms accounted for 76 percent of approved providers in May 2005 and served 59 percent of participating students in the previous school year (2003-04). A growing number and percentage of faith-based organizations have obtained state approval, rising from 18 providers (2 percent of providers) in May 2003 to 249 (9 percent) in May 2005, but they served less than one-half of one percent of student participants in 2003-04. School districts and public schools accounted for 17 percent of providers in May 2005, but served a larger proportion of participants (40 percent in 2003-04) (see Exhibit E-8).²³



2. Parental Notification

How and when do districts and schools inform parents of eligible children about the Title I school choice and supplemental services options?

The timing of parental notification was often too late to enable parents to choose a new school before the start of the 2004-05 school year. Almost half (49 percent) of districts notified parents after the school year had already started, and in these districts this notification occurred, on average, five weeks after the start of the school year.²⁴

3. Monitoring of Supplemental Service Providers

How are states monitoring and evaluating the effectiveness of supplemental service providers?

States report that they are working to develop and implement systems for monitoring and evaluating the performance of supplemental service providers, but, as of early 2005, 15 states had not established any monitoring process, 25 states had not yet established any standards for evaluating provider effectiveness, and none had finalized their evaluation standards. Seventeen states say they will evaluate student achievement on state assessments, although only one of these plans to use a matched control group. The most common approaches that states have implemented to monitor providers are surveying the

districts about provider effectiveness (25 states) and using providers' reports on student-level progress (18 states).²⁵

G. Teacher Quality and Professional Development

1. State Definitions of Highly Qualified Teachers

How have states implemented the requirements to define “highly qualified teacher” and to develop a “high objective uniform state standard of evaluation” (HOUSSE)?

Most states meet the requirement to test the content knowledge of new teachers through the Praxis II subject assessments developed by the Educational Testing Service (41 states). States vary considerably in the passing scores that they require teachers to obtain on the Praxis II exams in order to be certified to teach or to be deemed “highly qualified” under NCLB.²⁶

Nearly all states (47) allowed veteran teachers to demonstrate their subject-matter competency through a high objective uniform state standard of evaluation (HOUSSE), as of the spring of 2005. The most common type of HOUSSE option involved a point system wherein teachers were allowed to accumulate a state-determined number of points in order to earn a highly qualified status (29 states). Most states allowed points to be earned retroactively for such things as successful completion of certain college courses (28 states) or publishing articles and/or receiving teaching awards or honors (23 states). Four states allowed teachers to earn some points for evidence of improved student achievement. Twenty-six states allowed teachers to earn one-quarter or more of their HOUSSE points for a specified number of years of prior teaching experience in their subject(s). Eight states used their current, initial teacher certification systems as their official HOUSSE option; they reported that the certification requirements contained high standards of subject-area expertise.²⁷

2. Teachers' Highly Qualified Status

How many teachers meet the NCLB requirement to be “highly qualified”?

The large majority of teachers across the country have been designated as “highly qualified” under NCLB. According to state-reported data for 42 states, 86 percent of classes were taught by highly qualified teachers in 2003-04.²⁸ Principal and teacher reports for 2004-05 provide somewhat lower estimates of the percentage of classes taught by highly qualified teachers, but this is because a sizeable percentage did not know their “highly qualified” status. For example, 74 percent of teachers reported that they were considered highly qualified under NCLB, but 23 percent said they did not know their status and only 2 percent said they were not highly qualified.²⁹

Students in schools that have been identified for improvement were more likely to be taught by teachers who were not highly qualified than were students in non-identified schools. For example, only one percent of elementary teachers in non-identified schools said they were considered not highly qualified, compared with 5 percent in schools that were in the

first or second year of being identified for improvement, 8 percent in schools in corrective action, and 6 percent of schools in restructuring.³⁰

Schools with high concentrations of poor and minority students have more teachers who are considered not highly qualified than do other schools. In high-poverty schools, for example, 5 percent of elementary teachers and 12 percent of secondary English and math teachers reported in 2004-05 that they were considered not highly qualified under NCLB, compared with one percent in low-poverty elementary schools and 3 percent in low-poverty secondary schools.³¹

3. Professional Development

To what extent are teachers participating in professional development activities that are sustained, intensive, and focused on instruction?

Most teachers reported receiving some professional development in reading and math content and instructional strategies, but fewer than one-quarter of the teachers participated in such training for more than 24 hours over the 2003-04 school year and summer. For example, 90 percent of elementary teachers participated in at least one hour of professional development focused on instructional strategies for teaching reading, but only 20 percent participated for more than 24 hours over the 2003-04 school year and summer.³²

Teachers in high-poverty schools were more likely to participate in professional development focused on reading and mathematics than were teachers in low-poverty schools. For example, 53 percent of secondary English teachers in high-poverty schools reported participating in professional development focused on in-depth study of topics in reading or English compared with 36 percent of their colleagues in low-poverty schools.

4. Qualifications of Title I Paraprofessionals

How many paraprofessionals meet the NCLB qualifications requirements?

According to principal reports, 63 percent of Title I instructional aides had been determined to meet NCLB qualification requirements as of the 2004-05 school year. However, 87 percent of Title I instructional aides indicated that they had at least two years of college (and/or an associate's degree) or had passed a paraprofessional assessment. Nearly one-quarter (23 percent) of Title I instructional aides reported that, of the time that they spent tutoring or working with students in a classroom, a teacher was present only half or less of this time.³³

End Notes

¹ U.S. Department of Education, Budget Service.

² Jay Chambers, Joanne Lieberman, Tom Parrish, Daniel Kaleba, James Van Campen, and Stephanie Stullich (2000). *Study of Education Resources and Federal Funding: Final Report*. Washington, DC: U.S. Department of Education, Office of the Under Secretary, Planning and Evaluation Service. The estimate of the percentage of public schools receiving Title I funds was updated based on the number of Title I schools reported on Consolidated State Performance Reports for 2002-03 divided by the total number of public elementary and secondary schools in 2001-02 from the NCES Common Core of Data.

³ Beth Sinclair (forthcoming). *State ESEA Title I Participation Information for 2002-03: Final Summary Report*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, Policy and Program Studies Service.

⁴ Beth Sinclair (forthcoming). *State ESEA Title I Participation Information for 2002-03: Final Summary Report*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, Policy and Program Studies Service.

⁵ U.S. Department of Education, National Center for Education Statistics, unpublished data on averaged freshman graduation rates. U.S. Department of Education, Policy and Program Studies Service analysis of state-reported graduation rates from Consolidated State Performance Reports and State Education Agency websites. State-reported rates for 2003 or 2004 were used for 16 states where 2002 rates were not available.

⁶ Consolidated State Performance Reports, 2003-04.

⁷ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *Study of State Implementation of Accountability and Teacher Quality Under No Child Left Behind*.

⁸ Consolidated State Performance Reports, 2003-04.

⁹ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *Study of State Implementation of Accountability and Teacher Quality Under No Child Left Behind*.

¹⁰ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *Study of State Implementation of Accountability and Teacher Quality Under No Child Left Behind*.

¹¹ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *Study of State Implementation of Accountability and Teacher Quality Under No Child Left Behind*.

¹² U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *Study of State Implementation of Accountability and Teacher Quality Under No Child Left Behind*.

¹³ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *Study of State Implementation of Accountability and Teacher Quality Under No Child Left Behind*.

¹⁴ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *Study of State Implementation of Accountability and Teacher Quality Under No Child Left Behind*.

¹⁵ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *Study of State Implementation of Accountability and Teacher Quality under No Child Left Behind*.

¹⁶ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *National Longitudinal Study of No Child Left Behind*.

¹⁷ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *National Longitudinal Study of No Child Left Behind*.

¹⁸ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *Study of State Implementation of Accountability and Teacher Quality Under No Child Left Behind*.

¹⁹ Center on Education Policy (2005). *From the Capital to the Classroom: Year 3 of the No Child Left Behind Act*. Washington, DC: Center on Education Policy.

²⁰ Christine Padilla, Katrina Woodworth, Andrea Lash, Patrick M. Shields, and Katrina G. Laguarda (2005). *Evaluation of Title I Accountability Systems and School Improvement Efforts: Findings From 2002-03*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, Policy and Program Studies Service.

²¹ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *National Longitudinal Study of No Child Left Behind*.

²² U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *National Longitudinal Study of No Child Left Behind*.

²³ Policy and Program Studies Service monthly reviews of State Education Agency Websites, conducted by Westat from May 2003 through May 2005.

-
- ²⁴ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *National Longitudinal Study of No Child Left Behind*.
- ²⁵ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *Study of State Implementation of Accountability and Teacher Quality Under No Child Left Behind*.
- ²⁶ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *Study of State Implementation of Accountability and Teacher Quality Under No Child Left Behind*.
- ²⁷ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *Study of State Implementation of Accountability and Teacher Quality Under No Child Left Behind*.
- ²⁸ Consolidated State Performance Reports, 2003-04.
- ²⁹ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *National Longitudinal Study of No Child Left Behind*.
- ³⁰ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *National Longitudinal Study of No Child Left Behind*.
- ³¹ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *National Longitudinal Study of No Child Left Behind*.
- ³² U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *National Longitudinal Study of No Child Left Behind*.
- ³³ U.S. Department of Education, Policy and Program Studies Service, unpublished data from the *National Longitudinal Study of No Child Left Behind*.

Executive Summary of Volume II: Closing the Reading Gap

A report prepared for the Institute of Education Sciences
by the Corporation for the Advancement of Policy Evaluation

Joseph Torgesen, Florida Center for Reading Research

David Myers, Allen Schirm, Elizabeth Stuart, Sonya Vartivarian, & Wendy Mansfield
Mathematica Policy Research

Fran Stancavage, American Institutes for Research

Donna Durno and Rosanne Javorsky
Allegheny Intermediate Unit

Cynthia Haan, Haan Foundation

Evaluation Context

According to the National Assessment of Educational Progress (U.S. Department of Education 2003), nearly 4 in 10 fourth graders read below the basic level. Unfortunately, these literacy problems get worse as students advance through school and are exposed to progressively more complex concepts and courses. Historically, nearly three-quarters of these students never attain average levels of reading skill. While schools are often able to provide some literacy intervention, many lack the resources—teachers skilled in literacy development and appropriate learning materials—to help older students in elementary school reach grade level standards in reading.

The consequences of this problem are life changing. Young people entering high school in the bottom quartile of achievement are substantially more likely than students in the top quartile to drop out of school, setting in motion a host of negative social and economic outcomes for students and their families.

For their part, the nation's 16,000 school districts are spending hundreds of millions of dollars on often untested educational products and services developed by textbook publishers, commercial providers, and nonprofit organizations. Yet we know little about the effectiveness of these interventions. Which ones work best, and for whom? Under what conditions are they most effective? Do these programs have the potential to close the reading gap?

To help answer these questions, we initiated an evaluation of either parts or all of four widely used programs for elementary school students with reading problems. The programs are Corrective Reading, Failure Free Reading, Spell Read P.A.T., and Wilson Reading, all of which are expected to be more intensive and skillfully delivered than the programs typically provided in public schools.¹ The programs incorporate explicit and systematic instruction in the basic reading skills in which struggling readers are frequently deficient. Corrective Reading, Spell Read P.A.T., and Wilson Reading were implemented to provide word-level instruction, whereas Failure Free Reading focused on building reading comprehension and vocabulary in addition to word-level skills. Recent reports from small-scale research and clinical studies provide some evidence that the reading skills of students with severe reading difficulties in late elementary school can be substantially improved by providing, for a sustained period of time, the kinds of skillful, systematic, and explicit instruction that these programs offer (Torgesen 2005).

A. Evaluation Purpose and Design

Conducted just outside Pittsburgh, Pennsylvania, in the Allegheny Intermediate Unit (AIU), the evaluation is intended to explore the extent to which the four reading programs can affect both the word-level reading skills (phonemic decoding, fluency, accuracy) and reading comprehension of students in grades three and five who were identified as struggling readers by their teachers and by low test scores. Ultimately, it will provide educators with rigorous evidence of what could happen in terms of reading improvement if intensive, small-group reading programs like the ones in this study were introduced in many schools.

This study is a large-scale, longitudinal evaluation comprising two main elements. The first element of the evaluation is an impact study of the four interventions. This evaluation report is addressing three broad types of questions related to intervention impacts:

- What is the impact of being in any of the four remedial reading interventions, considered as a group, relative to the instruction provided by the schools? What is the impact of being in one of the remedial reading programs that focuses primarily on developing word-level skills, considered as a group, relative to the instruction provided by the schools? What is the impact of being in each of the four particular remedial reading interventions, considered individually, relative to the instruction provided by the schools?
- Do the impacts of programs vary across students with different baseline characteristics?
- To what extent can the instruction provided in this study close the reading gap and bring struggling readers within the normal range, relative to the instruction provided by their schools?

To answer these questions, the impact study was based on a scientifically rigorous design—an experimental design that uses random assignment at two levels: (1) 50 schools from 27 school districts were randomly assigned to one of the four interventions, and (2) within each school, eligible children in grades 3 and 5 were randomly assigned to a treatment group or to a control group. Students assigned to the intervention group (treatment group) were placed by the program providers and local coordinators into instructional groups of three students. Students in the control groups received the same instruction in reading that they would have ordinarily received. Children were defined as eligible if they were identified by their teachers as struggling readers and if they scored at or below the 30th percentile on a word-level reading test and at or above the 5th percentile on a vocabulary test. From an original pool of 1,576 3rd and 5th grade students identified as struggling readers, 1,042 also met the test-score criteria. Of these eligible students, 772 were given permission by their parents to participate in the evaluation.

The second element of the evaluation is an implementation study that has two components: (1) an exploration of the similarities and differences in reading instruction offered in the four interventions and (2) a description of the regular instruction that students in the control group received in the absence of the interventions and the regular instruction received by the treatment group beyond the interventions.

Test data and other information on students, parents, teachers, classrooms, and schools are being collected several times over a three-year period. Key data collection points pertinent to this summary report include the period just before the interventions began, when baseline information was collected, and the period immediately after the interventions ended, when follow-up data were collected. Additional follow-up data for students and teachers are being collected in 2005 and again in 2006.

B. The Interventions

We did not design new instructional programs for this evaluation. Rather, we employed either parts or all of four existing and widely used remedial reading instructional programs: Spell Read P.A.T., Corrective Reading, Wilson Reading, and Failure Free Reading.

As the evaluation was originally conceived, the four interventions would fall into two instructional classifications with two interventions in each. The interventions in one classification would focus only on word-level skills, and the interventions in the other classification would focus equally on word-level skills and reading comprehension/vocabulary.

Corrective Reading and Wilson Reading were modified to fit within the first of these classifications. The decision to modify these two intact programs was justified both because it created two treatment classes that were aligned with the different types of reading deficits observed in struggling readers and because it gave us sufficient statistical power to contrast the relative effectiveness of the two classes. Because Corrective Reading and Wilson Reading were modified, results from this study do not provide complete evaluations of these interventions; instead, the results suggest how interventions using primarily the word-level components of these programs will affect reading achievement.

With Corrective Reading and Wilson Reading focusing on word-level skills, it was expected that Spell Read P.A.T. and Failure Free Reading would focus on both word-level skills and reading comprehension/vocabulary. In a time-by-activity analysis of the instruction that was actually delivered, however, it was determined that three of the programs—Spell Read P.A.T., Corrective Reading, and Wilson Reading—focused primarily on the development of word-level skills, and one—Failure Free Reading—provided instruction in both word-level skills and the development of comprehension skills and vocabulary.

- ***Spell Read Phonological Auditory Training (P.A.T.)*** provides systematic and explicit fluency-oriented instruction in phonemic awareness and phonics along with every-day experiences in reading and writing for meaning. The phonemic activities include a wide variety of specific tasks focused on specific skill mastery and include, for example, building syllables from single sounds, blending consonant and vowel sounds, and analyzing or breaking syllables into their individual sounds. Each lesson also includes reading and writing activities intended to help students apply their phonically based reading skills to authentic reading and writing tasks. The Spell Read intervention had originally been one of the two “word-level plus comprehension” interventions, but after the time x activity analysis, we determined that it was more appropriately grouped as a “word-level” intervention.
- ***Corrective Reading*** uses scripted lessons that are designed to improve the efficiency of instruction and to maximize opportunities for students to respond and receive feedback. The lessons involve very explicit and systematic instructional sequences, including a series of quick tasks that are intended to focus students’ attention on critical elements for successful word identification as well as exercises intended to build rate and fluency through oral reading of stories that have been constructed to counter word-guessing habits. Although the Corrective Reading program does have instructional procedures that focus on comprehension, they were originally designated as a “word-level intervention,” and the developer was asked not to include these elements in this study.
- ***Wilson Reading*** uses direct, multi-sensory, structured teaching based on the Orton-Gillingham methodology. The program is based on 10 principles of

instruction, some of which involve teaching fluent identification of letter sounds; presenting the structure of language in a systematic, cumulative manner; presenting concepts in the context of controlled as well as non-controlled text; and teaching and reinforcing concepts with visual-auditory-kinesthetic-tactile methods. Similar to Corrective Reading, the Wilson Program has instructional procedures that focus on comprehension and vocabulary, but since they were originally designated as a “word-level” intervention, they were asked not to include these in this study.

- ***Failure Free Reading*** uses a combination of computer-based lessons, workbook exercises, and teacher-led instruction to teach sight vocabulary, fluency, and comprehension. The program is designed to have students spend approximately one-third of each instructional session working within each of these formats, so that they are not taught simultaneously as a group. Unlike the other three interventions in this study, Failure Free does not emphasize phonemic decoding strategies. Rather, the intervention depends upon building the student’s vocabulary of “sight words” through a program involving multiple exposures and text that is engineered to support learning of new words. Students read material that is designed to be of interest to their age level while also challenging their current independent and instructional reading level. Lessons are based on story text that is controlled for syntax and semantic content.

C. Measures of Reading Ability

Seven measures of reading skill were administered at the beginning and end of the school year to assess student progress in learning to read. As outlined below, these measures of reading skills assessed phonemic decoding, word reading accuracy, text reading fluency, and reading comprehension.

Phonemic Decoding

- Word Attack (WA) subtest from the Woodcock Reading Mastery Test-Revised (WRMT-R)
- Phonemic Decoding Efficiency (PDE) subtest from the Test of Word Reading Efficiency (TOWRE)

Word Reading Accuracy and Fluency

- Word Identification (WI) subtest from the WRMT-R
- Sight Word Efficiency (SWE) subtest from the TOWRE
- Oral Reading Fluency subtest from Edformation, Inc. The text of this report refers to the reading passages as “AimswEB” passages, which is the term used broadly in the reading practice community.

Reading Comprehension

- Passage Comprehension (PC) subtest from the WRMT-R
- Passage Comprehension from the Group Reading Assessment and Diagnostic Evaluation (GRADE)

For all tests except the Aimsweb passages, the analysis uses grade-normalized standard scores, which indicate where a student falls within the overall distribution of reading ability among students in the same grade. Scores above 100 indicate above-average performance; scores below 100 indicate below-average performance. In the population of students across the country at all levels of reading ability, standard scores are constructed to have a mean of 100 and a standard deviation of 15, implying that approximately 70 percent of all students' scores will fall between 85 and 115 and that approximately 95 percent of all students' scores will fall between 70 and 130. For the Aimsweb passages, the score used in this analysis is the median correct words per minute from three grade-level passages.

D. Implementing the Interventions

The interventions were implemented from the first week of November 2003 through the first weeks in May 2004. During this time students received, on average, about 90 hours of instruction, which was delivered five days a week to groups of three students in sessions that were approximately 50 minutes long. A small part of the instruction was delivered in groups of two, or 1:1, because of absences and make-up sessions. Since many of the sessions took place during the student's regular classroom reading instruction, teachers reported that students in the treatment groups received less reading instruction in the classroom than did students in the control group (1.2 hours per week versus 4.4 hours per week.). Students in the treatment group received more small-group instruction than did students in the control group (6.8 hours per week versus 3.7 hours per week). Both groups received a very small amount of 1:1 tutoring in reading from their schools during the week.

Teachers were recruited from participating schools on the basis of experience and the personal characteristics relevant to teaching struggling readers. They received, on average, nearly 70 hours of professional development and support during the implementation year as follows:

- About 30 hours during an initial week of intensive introduction to each program
- About 24 hours during a seven-week period at the beginning of the year when the teachers practiced their assigned methods with 4th-grade struggling readers in their schools
- About 14 hours of supervision during the intervention phase

According to an examination of videotaped teaching sessions by the research team, the training and supervision produced instruction that was judged to be faithful to each intervention model. The program providers themselves also rated the teachers as generally above average in both

their teaching skill and fidelity to program requirements relative to other teachers with the same level of training and experience.

E. Characteristics of Students in the Evaluation

The characteristics of the students in the evaluation sample are shown in Table 1 (see the end of this summary for all tables). About 45 percent of the students qualified for free or reduced-price lunches. In addition, about 27 percent were African American, and 73 percent were white. Fewer than two percent were Hispanic. Roughly 33 percent of the students had a learning disability or other disability.

On average, the students in our evaluation sample scored about one-half to one standard deviation below national norms (mean 100 and standard deviation 15) on measures used to assess their ability to decode words. For example, on the Word Attack subtest of the Woodcock Reading Mastery Test-Revised (WRMT-R), the average standard score was 93. This translates into a percentile ranking of 32. On the TOWRE test for phonemic decoding efficiency (PDE), the average standard score was 83, at approximately the 13th percentile. On the measure of word reading accuracy (Word Identification subtest for the WRMT-R), the average score placed these students at the 23rd percentile. For word reading fluency, the average score placed them at the 16th percentile for word reading efficiency (TOWRE SWE), and third- and fifth-grade students, respectively, read 41 and 77 words per minute on the oral reading fluency passages (Aimsweb). In terms of reading comprehension, the average score for the WRMT-R test of passage comprehension placed students at the 30th percentile, and for the Group Reading and Diagnostic Assessment (GRADE), they scored, on average, at the 23rd percentile.

This sample, as a whole, was substantially less impaired in basic reading skills than most samples used in previous research with older reading disabled students. These earlier studies typically examined samples in which the phonemic decoding and word reading accuracy skills of the average student were below the tenth percentile and, in some studies, at only about the first or second percentile. Students in such samples are much more impaired and more homogeneous in their reading abilities than the students in this evaluation and in the population of all struggling readers in the United States. Thus, it is not known whether the findings from these previous studies pertain to broader groups of struggling readers in which the average student's reading abilities fall between, say, the 20th and 30th percentiles. This evaluation can help to address this issue. It obtained a broad sample of struggling readers, and is evaluating in regular school settings the kinds of intensive reading interventions that have been widely marketed by providers and widely sought by school districts to improve such students' reading skills.

F. Discussion of Impacts

This first year report assesses the impact of the four interventions on the treatment groups in comparison with the control groups immediately after the end of the reading interventions. In particular, we provide detailed estimates of the impacts, including the impact of being randomly assigned to receive any of the interventions, being randomly assigned to receive a word-level intervention, and being randomly assigned to receive each of the individual interventions. For

purposes of this summary, we focus on the impact of being randomly assigned to receive any intervention compared to receiving the instruction that would normally be provided. These findings are the most robust because of the larger sample sizes. The full report also estimates impacts for various subgroups, including students with weak and strong initial word attack skills, students with low or high beginning vocabulary scores, and students who either qualified or did not qualify for free or reduced price school lunches.²

The impact of each of the four interventions is the difference between average treatment and control group outcomes. Because students were randomly assigned to the two groups, we would expect the groups to be statistically equivalent; thus, with a high probability, any differences in outcomes can be attributed to the interventions. Also because of random assignment, the outcomes themselves can be defined either as test scores at the end of the school year, or as the change in test scores between the beginning and end of the school year (the “gain”). In the tables of impacts (Tables 2-4), we show three types of numbers. The baseline score shows the average standard score for students at the beginning of the school year. The control gain indicates the improvement that students would have made in the absence of the interventions. Finally, the impact shows the value added by the interventions. In other words, the impact is the amount that the interventions increased students’ test scores relative to the control group. The gain in the intervention group students’ average test scores between the beginning and end of the school year can be calculated by adding the control group gain and the impact.

In practice, impacts were estimated using a hierarchical linear model that included a student-level model and a school-level model. In the student-level model, we include indicators for treatment status and grade level as well as the baseline test score. The baseline test score was included to increase the precision with which we measured the impact, that is, to reduce the standard error of the estimated impact. The school-level model included indicators that show the intervention to which each school was randomly assigned and indicators for the blocking strata used in the random assignment of schools to interventions. Below, we describe some of the key interim findings:

- ***For third graders, we found that the four interventions combined had impacts on phonemic decoding, word reading accuracy and fluency, and reading comprehension. There are fewer significant impacts for fifth graders than for third graders (see Table 2). The impacts of the three word-level interventions combined were similar to those for all four interventions combined.*** Although many of the impacts shown in Table 2 for third graders are positive and statistically significant when all, or just the three word-level, interventions are considered, it is noteworthy that on the GRADE, which is a group-administered test for reading comprehension, the impact estimate and the estimated change in standard scores for the control group indicate that there was not a substantial improvement in reading comprehension in the intervention groups relative to the larger normative sample for the test. Instead, this evidence suggests that the interventions helped these students maintain their relative position among all students and not lose ground in reading comprehension, as measured by the GRADE test. Results from the GRADE test are particularly important, because this test, more than others in the battery, closely mimics the kinds of testing demands (group administration,

responding to multiple choice comprehension questions) found in current state-administered reading accountability measures.

- ***Among key subgroups, the most notable variability in findings was observed for students who qualified for free or reduced price lunches and those who did not.*** Although the ability to compare impacts between groups is limited by the relatively small samples, we did generally find significant impacts on the reading outcomes for third graders who did not qualify and few significant impacts for those who did qualify (see Tables 3 and 4), when all four interventions are considered together and when the three word-level interventions are considered together. These findings for third graders may be driven in part by particularly large negative gains among the control group students in the schools assigned to one intervention.
- ***At the end of the first year, the reading gap for students in the intervention group was generally smaller than the gap for students in the control group when considering all four interventions together.*** The reading gap describes the extent to which the average student in one of the two evaluation groups (intervention or control) is lagging behind the average student in the population (see Figures 1-12 and Table 5). The reduction in the reading gap attributable to the interventions at the end of the school year is measured by the interventions' impact relative to the gap for the control group, the latter showing how well students would have performed if they had not been in one of the interventions. Being in one of the interventions reduced the reading gap on Word Attack skills by about two-thirds for third graders. On other word-level tests and a measure of reading comprehension, the interventions reduced the gap for third graders by about one-fifth to one-quarter. For fifth graders, the interventions reduced the gap for Word Attack and Sight Word Efficiency by about 60 and 12 percent, respectively.³

Future reports will focus on the impacts of the interventions one year after they ended. At this point, it is still too early to draw definitive conclusions about the impact of the interventions assessed in this study. Based on the results from earlier research (Torgesen et al. 2001), there is a reasonable possibility that students who substantially improved their phonemic decoding skills will continue to improve in reading comprehension relative to average readers. Consistent with the overall pattern of immediate impacts, we would expect more improvement in students who were third graders when they received the intervention relative to fifth graders. We are currently processing second-year data (which includes scores on the Pennsylvania state assessments) and expect to release a report on that analysis within the next year.

End Notes

¹ These four interventions were selected from more than a dozen potential program providers by members of the Scientific Advisory Board of the Haan Foundation for Children. See Appendix Q in Volume II for a list of the Scientific Advisory Board members.

² The impacts described here represent the impact of being selected to participate in one of the interventions. A small number of students selected for the interventions did not participate, and about 7.5 percent received less than a full dose (80 hours) of instruction. Estimation of the effect of an intervention on participants and

those who participated for 80 or more hours requires that stronger assumptions be made than when estimating impacts for those offered the opportunity to participate, and we cannot have the same confidence in the findings as we do with the results discussed in this summary. Our full report presents estimates of the effects for participants and those who participated for at least 80 hours. These findings are similar to those reported here.

³ In future analyses, we plan to explore another approach for estimating the impact of the interventions on closing the reading gap. This approach will contrast the percentage of students in the intervention groups and the control groups who scored within the “normal range” on the standardized tests.

Bibliography

Torgesen, J.K. “Recent Discoveries from Research on Remedial Interventions for Children with Dyslexia.” In M. Snowling and C. Hulme, eds., *The Science of Reading*. Oxford: Blackwell Publishers, 2005.

Torgesen, J.K., A.W. Alexander, R.K. Wagner, C.A. Rashotte, K. Voeller, T. Conway, and E. Rose. “Intensive Remedial Instruction for Children with Severe Reading Disabilities: Immediate and Long-Term Outcomes from Two Instruction Approaches.” *Journal of Learning Disabilities* 2001, 34: 33-58.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. “The Nation’s Report Card: Reading Highlights 2003.” NCES 2004-452. Washington, DC: NCES. Available online at [www.nces.ed.gov/nationsreportcard/pdf/main2003/2004452.pdf].

Table 1
Baseline Characteristics of the Analysis Sample
3rd Grade and 5th Grade

Baseline Means	Grade Level					
	Combined		3rd		5th	
Student Characteristics						
Age	9.7		8.7		10.7	
Male (%)	54		52		56	
Hispanic (%)	2		2		1	
Race--White (%)	73		71		74	
Race--African American (%)	27		29		26	
Race--Other (%)	a		a		a	
Family income less than \$30,000 (%)	50		49		50	
Family income between \$30,000 and \$60,000 (%)	34		33		35	
Family income over \$60,000 (%)	16		18		14	
Eligible for Free or Reduced Price Lunch (%)	45		46		45	
Has any learning or other disability (%)	33		34		32	
Mother has bachelor's degree or higher (%)	12		12		12	
	Standard		Standard		Standard	
Reading Tests	Score	Percentile	Score	Percentile	Score	Percentile
Screening Tests						
TOWRE Sight Word Efficiency	84.3	15	84.4	15	84.2	15
TOWRE Phonemic Decoding Efficiency	82.9	13	85.6	17	80.5	10
Peabody Picture Vocabulary Test--Revised	94.8	36	94.6	36	94.9	37
Baseline Tests						
WRM Word Identification	88.7	23	88.7	23	88.7	22
TOWRE Phonemic Decoding Efficiency	83.2	13	85.6	17	81.0	10
WRM Word Attack	92.9	32	92.6	31	93.1	32
TOWRE Sight Word Efficiency	85.3	16	86.5	18	84.2	15
AIMSWeb (Raw score)	NA	NA	40.9	NA	77.4	NA
WRM Passage Comprehension	92.3	30	91.8	29	92.7	31
GRADE	89.0	23	86.3	18	91.4	28
Woodcock Johnson Spelling	89.7	25	88.6	22	90.8	27
Woodcock Johnson Calculation	94.9	37	95.4	38	94.6	36
Other Baseline Tests Administered						
RAN Colors	89.0	23	87.7	21	90.2	26
RAN Letters	89.7	25	87.0	19	92.1	30
RAN Numbers	92.0	30	89.6	24	94.3	35
RAN Objects	88.8	23	87.7	21	89.8	25
RAS Numbers and Letters	89.3	24	87.1	19	91.4	28
RAS Colors, Numbers, and Letters	88.9	23	86.6	19	91.0	27
CTOPP Blending Words	7.5	20	7.7	22	7.3	18
CTOPP Elision	7.7	22	7.9	25	7.5	20
CTOPP Rapid Digit Naming	7.9	24	7.8	24	8.0	25
CTOPP Rapid Letter Naming	8.5	30	8.5	31	8.4	30
Clinical Evaluation of Language Fundamentals-IV	7.8	23	7.6	21	8.0	25
Sample Size	742		335		407	

Note: Weights used to account for differential randomization probabilities and nonresponse.

Note: All standard scores have mean 100 and standard deviation 15, except for CTOPP and Clinical Evaluation of Language Fundamentals-IV, which have mean 10 and standard deviation 3. Standard scores unavailable for the Aimsweb test.

Note: The percentile score shown for each test is the percentile corresponding with the mean standard score.

a Values suppressed to protect student confidentiality.

Table 2
Impacts for 3rd and 5th Graders

	Baseline	All Interventions		Word-level interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
		Control	ABCD	Control	BCD	Control	A	Control	B	Control	C	Control	D
Grade 3		Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact
Word Attack	92.6	0.2	5.0 *	0.0	6.8 *	0.7	-0.5	2.5	6.5 *	-3.0	8.8 *	0.5	5.2 *
TOWRE PDE	85.6	3.0	3.0 *	2.6	4.4 *	4.1	-1.3	4.1	7.1 *	0.2	5.8 *	3.6	0.4
Word Identification	88.7	-0.6	2.3 *	-0.6	2.6 *	-0.5	1.3	0.4	2.0	-2.3	2.5	0.1	3.3 *
TOWRE SWE	86.5	3.4	2.7 *	3.6	2.8 *	2.9	2.6	4.9	0.7	3.5	3.1	2.4	4.6 *
Aimsweb	40.9	20.6	4.9 *	20.3	5.9 *	21.5	1.9	22.6	1.0	17.5	6.0	20.9	10.7 *
Passage Comprehension	91.8	0.9	1.2	1.5	0.7	-0.8	2.7	2.4	0.2	-0.5	1.0	2.6	0.9
GRADE	86.2	-4.0	4.6 *	-3.1	4.4	-6.5	5.3	-4.2	4.9	-4.3	4.2	-0.9	4.2
Sample Size	335		335		242		93		92		71		79
Grade 5		Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact
Word Attack	93.1	2.2	2.7 *	2.4	3.9 *	1.3	-0.9	3.2	5.3 *	2.0	4.4 *	2.1	1.9
TOWRE PDE	81.0	5.9	1.4	6.3	1.5	4.6	1.1	7.9	4.1 *	6.8	-1.4 #	4.3	1.9
Word Identification	88.7	2.9	0.5	2.8	0.9	3.1	-0.6	2.8	0.1	2.6	2.1	3.1	0.3
TOWRE SWE	84.2	4.0	1.4 *	4.5	1.3	2.4	1.7	5.6	2.1	4.6	-0.5	3.4	2.2
Aimsweb	77.4	19.1	2.0	18.7	2.8	20.5	-0.3	19.6	3.6	19.4	-0.1	17.1	4.9
Passage Comprehension	92.7	-1.7	1.3	-2.1	1.6	-0.6	0.3	-1.2	0.6	-3.7	2.5	-1.4	1.8
GRADE	91.5	1.0	-0.2	0.8	0.3	1.6	-1.6	-0.5	-0.7	-0.7	1.3	3.6	0.3
Sample Size	407		407		281		126		104		91		86

* Impact is statistically significant at the 0.05 level.

Impact is statistically different from the 3rd grade impact at the 0.05 level.

Note: Sample sizes indicate the number of students randomly assigned to the intervention or control group, excluding students with missing test scores at the beginning or end of the school year.

Table 3
Impacts for 3rd and 5th Graders Eligible for Free or Reduced Price School Lunch

Grade 3	Baseline	All Interventions		Word-level interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading				
		Control	ABCD	Control	BCD	Control	A	Control	B	Control	C	Control	D			
Word Attack	92.2	1.3	4.7 *	1.6	5.9 *	0.7	1.3	1.7	8.4 *	0.2	6.0 *	#	2.8	3.3		
TOWRE PDE	85.3	4.6	1.8	4.5	2.6	#	4.9	-0.7	5.1	6.2 *	1.9	3.6	#	6.5	-2.0	
Word Identification	88.0	0.2	1.1	0.3	1.1		-0.2	1.0	2.3	-0.6	-1.4	1.2		0.0	2.8	
TOWRE SWE	85.5	3.5	1.3	4.0	0.7		2.2	3.0	4.1	-0.8	3.9	2.5		3.9	0.4	#
Aimswab	38.6	20.3	2.0	19.6	3.1		22.5	-1.1	22.0	-1.9	16.1	6.4		20.7	4.7	
Passage Comprehension	90.4	3.3	-0.8	#	4.2	-1.2	#	0.7	0.4	3.5	0.5	4.5	-2.6	#	4.5	-1.5
GRADE	84.4	-2.0	0.1	#	-0.7	-0.8	#	-6.0	2.5	-2.6	1.6	-1.4	-2.1	#	1.8	-1.7
Sample Size	193															
Grade 5	Baseline	All Interventions		Word-level interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading				
		Control	ABCD	Control	BCD	Control	A	Control	B	Control	C	Control	D			
Word Attack	92.5	3.5	0.6	4.1	1.5	1.7	-2.3	5.7	0.8	#	3.7	3.0	2.8	0.8		
TOWRE PDE	80.1	6.5	0.6	6.6	1.0	6.2	-0.5	8.9	2.9		7.2	-1.2	3.8	1.3		
Word Identification	87.8	2.4	0.4	2.5	0.7	2.2	-0.4	2.5	-1.2		2.1	3.0 *	3.0	0.2		
TOWRE SWE	83.2	2.6	3.7 *	#	2.9	3.8 *	#	1.6	3.2	4.5	3.9 *	4.1	1.0	0.3	6.5 *	#
Aimswab	73.4	14.7	3.1	14.0	4.5	16.6	-1.1	16.0	8.6 *		13.7	0.7	12.4	4.4		
Passage Comprehension	90.6	-0.1	-0.3	-0.3	-0.1	0.5	-0.9	0.6	-0.8		-0.8	-0.8	-0.8	1.3		
GRADE	88.6	3.2	-4.1 *	#	3.1	-3.7		3.3	-5.4	4.9	-6.1 *	1.0	-4.2	3.3	-0.8	
Sample Size	230															

* Impact is statistically significant at the 0.05 level.

Impact is statistically different from the impact for all students in that grade at the 0.05 level.

Table 4
Impacts for 3rd and 5th Graders Not Eligible for Free or Reduced Price School Lunch

	Baseline	All Interventions		Word-level interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
		Control Gain	ABCD Impact	Control Gain	BCD Impact	Control Gain	A Impact	Control Gain	B Impact	Control Gain	C Impact	Control Gain	D Impact
Grade 3													
Word Attack	93.3	-2.7	7.8 *	-3.8	10.9 *	0.7	-1.7	0.8	8.3 *	-13.2	19.5 * #	0.9	5.0
TOWRE PDE	86.1	0.1	5.3 *	-1.2	8.0 * #	4.1	-3.1	4.8	6.2 *	-12.1	17.6 * #	3.7	0.3
Word Identification	89.9	-2.4	3.6 *	-3.1	4.6 *	-0.2	0.5	-1.1	2.4	-7.8	7.8	-0.3	3.6
TOWRE SWE	87.9	3.0	3.0 *	2.6	3.9 *	4.1	0.2	6.8	-0.5	-0.1	5.2	1.1	6.9 * #
Aimsweb	44.1	19.0	7.6 *	19.0	8.4 *	19.1	5.1	23.1	1.1	13.0	9.6	20.9	14.5 *
Passage Comprehension	93.8	-5.0	6.1 * #	-5.9	6.7 * #	-2.1	4.2	2.7	-2.8	-20.9	19.5 * #	0.5	3.6
GRADE	88.9	-8.6	9.5 * #	-8.9	10.6 * #	-7.5	6.4	-5.5	6.0	-17.9	19.2 * #	-3.4	6.6
Sample Size	142												
Grade 5													
Word Attack	94.0	1.4	3.7 *	1.5	5.1 *	0.9	-0.5	1.3	8.9 * #	1.4	4.1	1.9	2.2
TOWRE PDE	82.0	5.3	1.2	6.1	1.0	3.0	1.6	6.3	4.8 *	6.9	-2.1	5.0	0.5
Word Identification	89.7	3.6	0.0	3.1	0.5	4.8	-1.6	2.5	0.9	3.8	0.5	3.1	0.0
TOWRE SWE	85.4	4.8	0.0 #	5.7	-0.7 #	1.9	2.0	5.3	1.1	5.0	-0.4	6.8	-2.8 #
Aimsweb	82.2	22.1	0.3	21.7	0.2	23.5	0.5	21.0	-0.7	22.0	0.0	22.0	1.4
Passage Comprehension	95.1	-2.9	2.1	-3.2	2.4	-1.9	1.4	-2.4	1.3	-6.9	5.3 *	-0.3	0.5
GRADE	94.9	0.3	1.2 #	-0.2	1.9	1.9	-0.7	-4.5	1.8	0.1	2.8	3.8	1.0
Sample Size	177												

* Impact is statistically significant at the 0.05 level.

Impact is statistically different from the impact for all students in that grade at the 0.05 level.

Table 5
Relative Gap Reduction: All Interventions Combined

	Average at baseline	Gap at baseline (Std. Units)	Average at follow-up		Gap at follow-up (Std. Units)		Impact	RGR
			Intervention Group	Control Group	Intervention Group	Control Group		
3rd Grade								
Word Attack	92.6	0.49	97.8	92.8	0.15	0.48	5.0 *	0.69
TOWRE PDE	85.6	0.96	91.6	88.6	0.56	0.76	3.0 *	0.26
Word Identification	88.7	0.75	90.4	88.1	0.64	0.79	2.3 *	0.19
TOWRE SWE	86.5	0.90	92.6	89.9	0.49	0.67	2.7 *	0.27
Aimsweb	NA	NA	NA	NA	NA	NA	NA	NA
Passage Comprehension	91.8	0.55	93.9	92.7	0.40	0.48	1.2	0.17
GRADE	86.2	0.92	86.9	82.3	0.87	1.18	4.6 *	0.26
5th Grade								
Word Attack	93.1	0.46	98.0	95.3	0.14	0.31	2.7 *	0.56
TOWRE PDE	81.0	1.27	88.3	86.9	0.78	0.87	1.4	0.11
Word Identification	88.7	0.76	92.1	91.6	0.53	0.56	0.5	0.06
TOWRE SWE	84.2	1.05	89.6	88.2	0.69	0.78	1.4 *	0.12
Aimsweb	NA	NA	NA	NA	NA	NA	NA	NA
Passage Comprehension	92.7	0.49	92.2	90.9	0.52	0.60	1.3	0.14
GRADE	91.5	0.57	92.3	92.5	0.51	0.50	-0.2	-0.02

* Impact is statistically significant at the 0.05 level.

Note: RGR defined as $RGR = (Impact / (100 - \text{Average for Control Group at follow-up}))$.

Note: Gap defined as $(100 - \text{Average Score}) / 15$, where 100 is the population average and 15 is the population standard deviation.

Note: Values for Aimsweb not available because normed standard scores were unavailable.

Figure 1

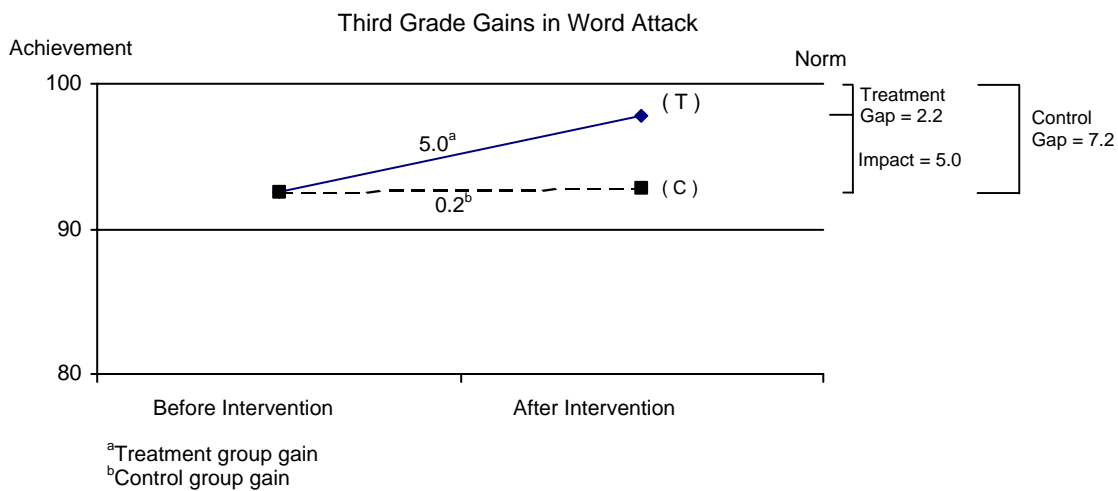


Figure 2

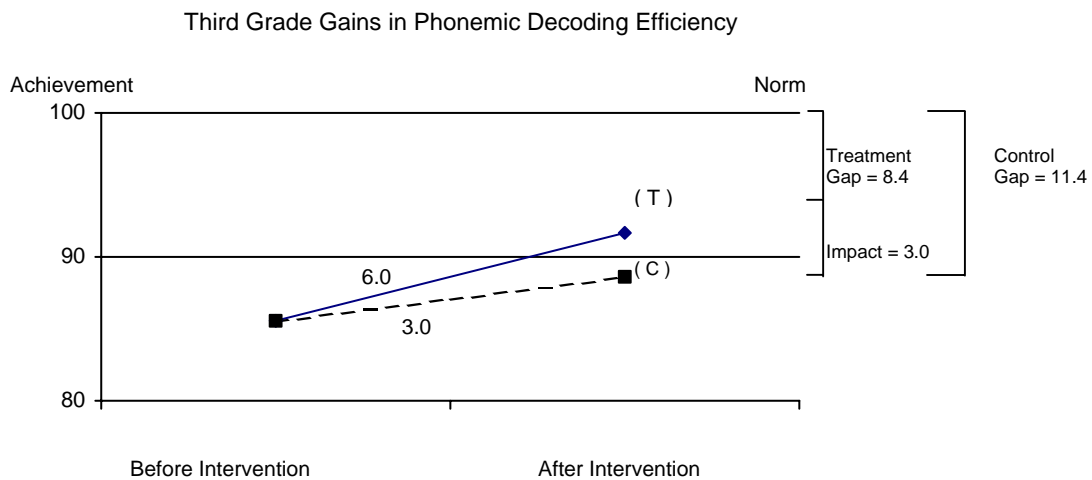


Figure 3

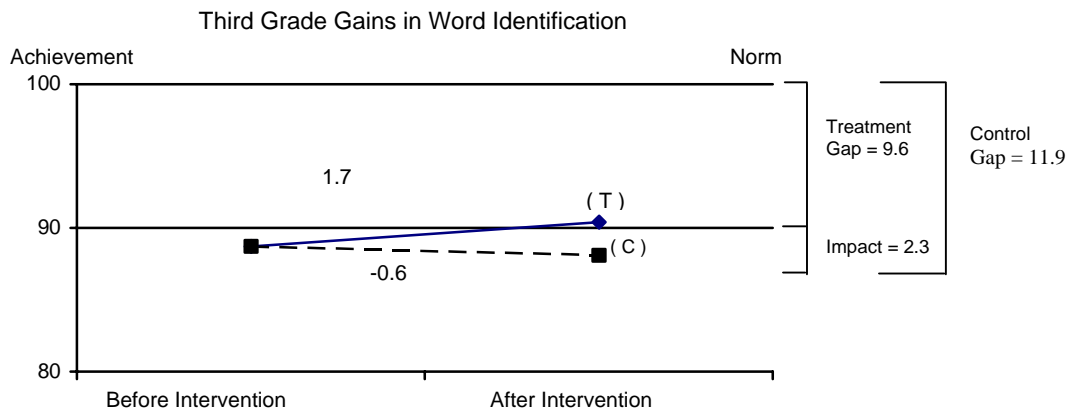


Figure 4

Third Grade Gains in Sight Word Efficiency

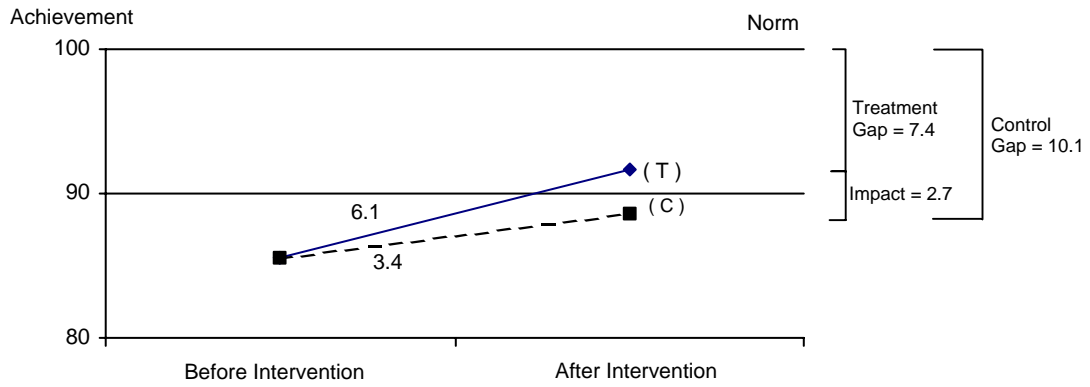


Figure 5

Third Grade Gains in Passage Comprehension

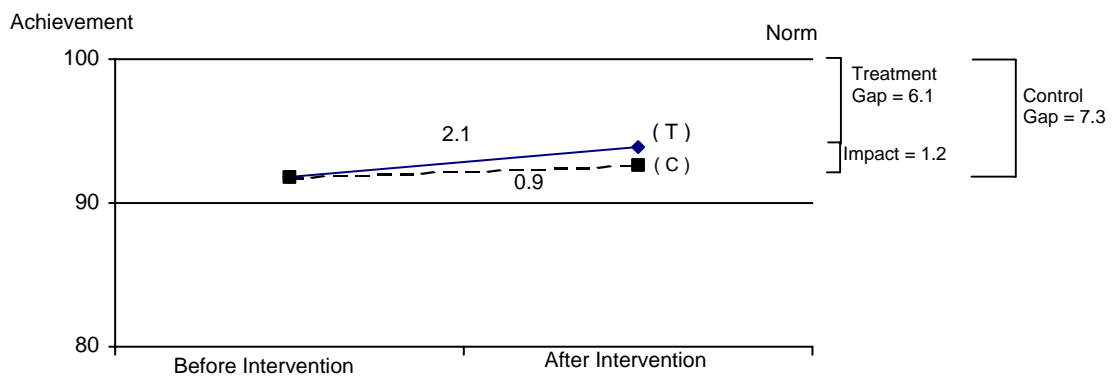


Figure 6

Third Grade Gains in GRADE Test

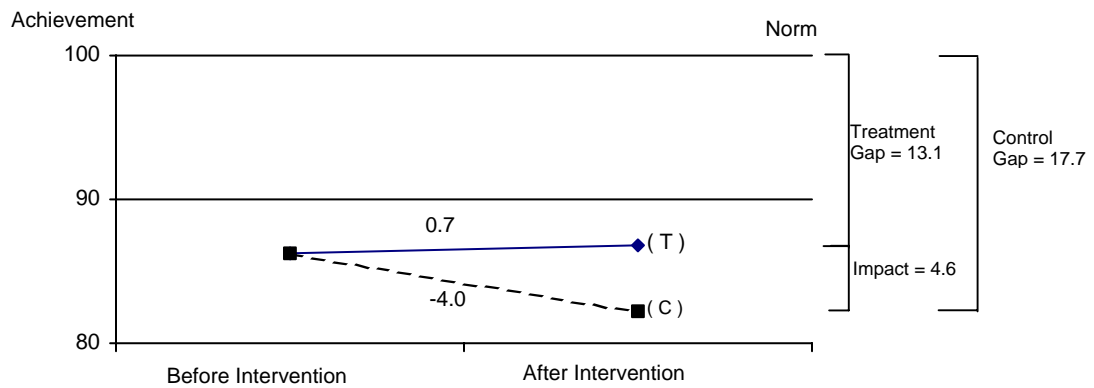


Figure 7

Fifth-Grade Gains in Word Attack

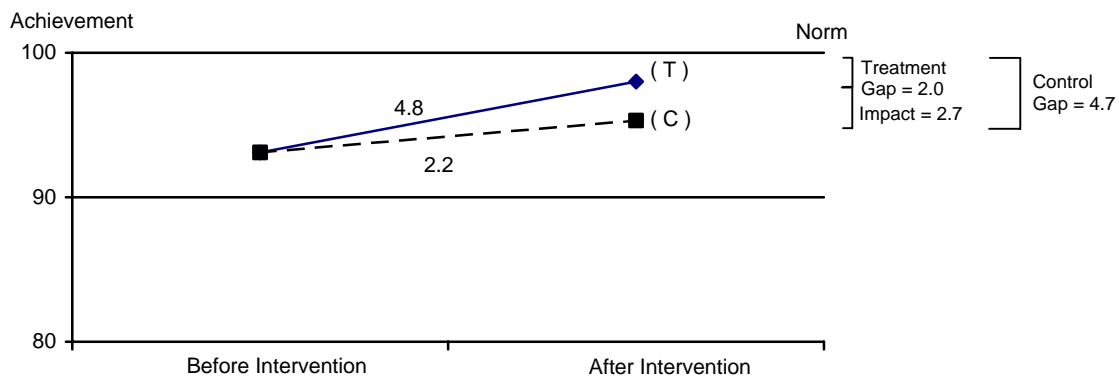


Figure 8

Fifth-Grade Gains in Phonemic Decoding Efficiency

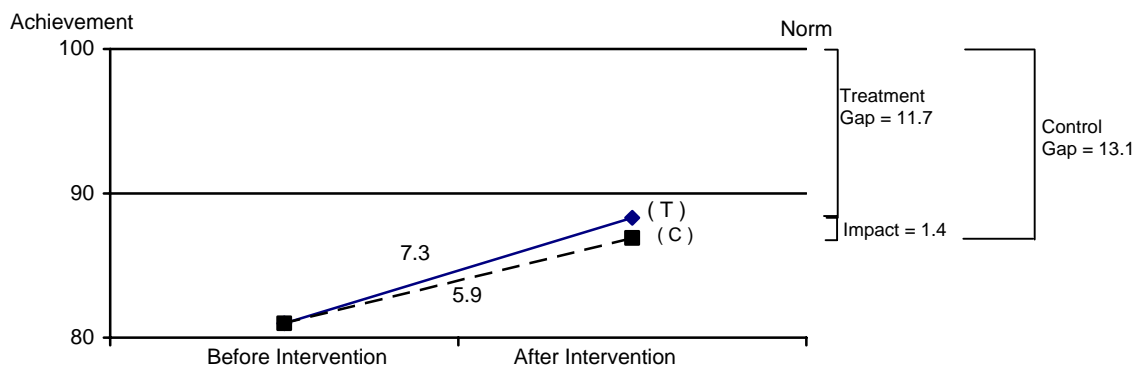


Figure 9

Fifth-Grade Gains in Word Identification

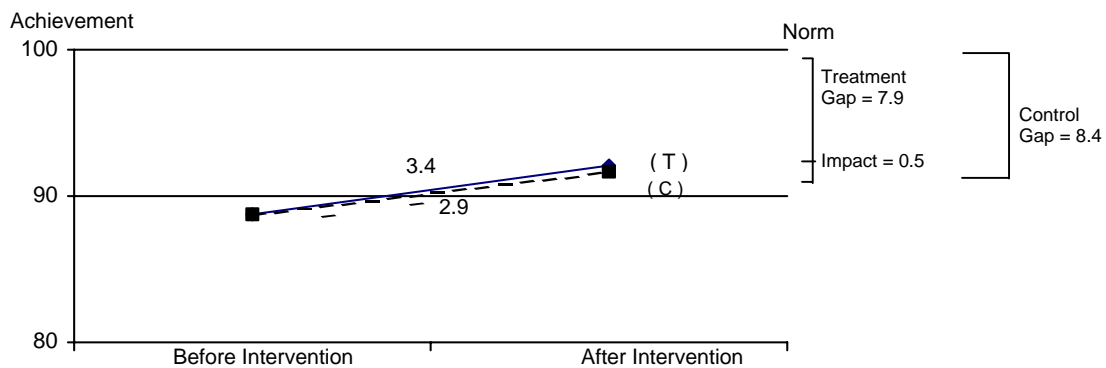


Figure 10

Fifth-Grade Gains in Sight Word Efficiency

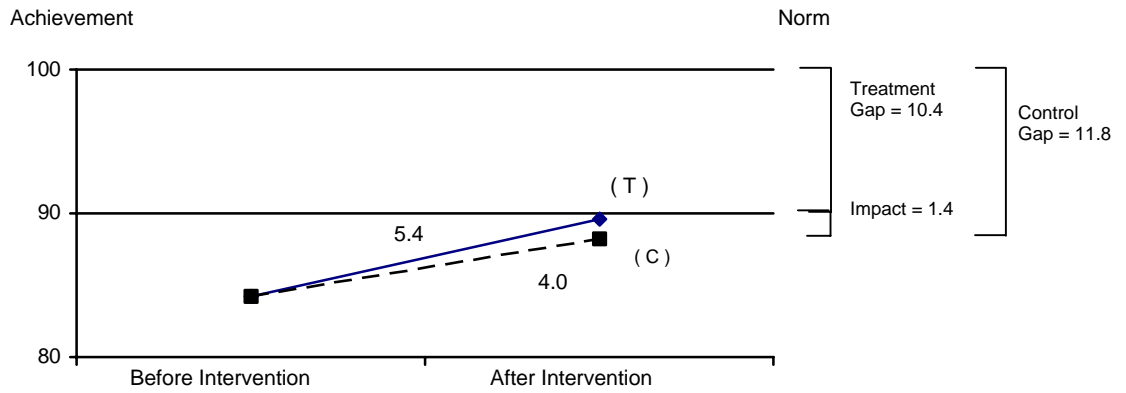


Figure 11

Fifth-Grade Gains in Passage Comprehension

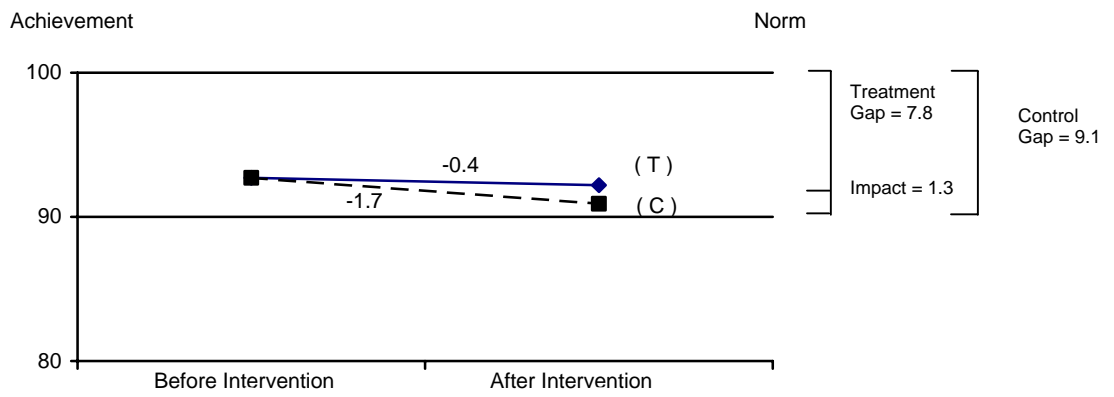
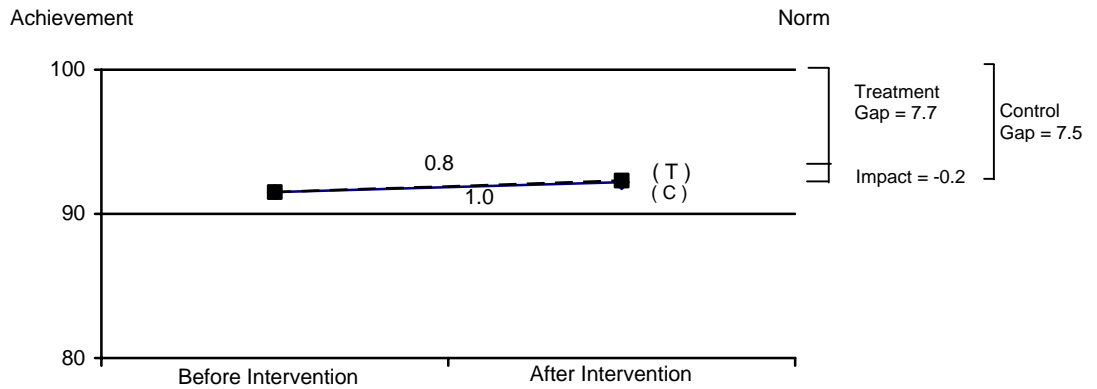


Figure 12

Fifth-Grade Gains in GRADE Test



Appendix A

Independent Review Panel Members

Kaleem Caire, American Education Reform Council

Tom Cook, Northwestern University

Chris Cross, Cross & Jofus, LLC

Gayle Fallon, Houston Federation of Teachers

David Francis, University of Houston

Norma Garza

Eric Hanushek, Hoover Institution, Stanford University

Sharon Johnson, Withrow University High School

Paul Peterson, Harvard University

Steve Raudenbush, University of Chicago

Eric Smith, Anne Arundel County Public Schools

John Stevens, Texas Business and Education Coalition

Patricia Supple, Archdiocese of Los Angeles

Tasha Tillman

Maris Vinovskis, University of Michigan

Rodney Watson, Louisiana Department of Education

