CORPORATION FOR THE ADVANCEMENT
OF POLICY EVALUATION

Closing the Reading Gap:
First Year Findings from
a Randomized Trial
of Four Reading
Interventions for Striving
Readers

Executive Summary

February 2006

Joseph Torgesen,
Florida Center for Reading Research

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

Fran Stancavage,
American Institutes for Research

Donna Durno and Rosanne Javorsky,
Allegheny Intermediate Unit

Cinthia Haan,
Haan Foundation

ABOUT CAPE

The Corporation for the Advancement of Policy Evaluation (CAPE) is a nonprofit organization that facilitates, conducts, and supports research activities to advance the development of public policy. CAPE also assists in determining the effectiveness of programs and policies designed to address social problems and informs the policy and service communities about its findings. Its major clients include private foundations and government agencies.

ACKNOWLEDGMENTS

EXECUTIVE SUMMARY

EVALUATION CONTEXT

According to the National Assessment of Education al 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).

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

---

[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 for a list of the Scientific Advisory Board members.

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 is 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.

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.

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.

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.

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.

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

_____

[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.

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 were 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.

---

[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.

Table 1

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

	Grade Level					
Baseline Means	Combined		3rd		5th	

Student Characteristics

	Combined	3rd	5th
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

Reading Tests	Standard Score	Percentile	Standard Score	Percentile	Standard 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 confidentialiy.

Table 2

Impacts for 3rd and 5th Graders

Grade 3

Grade 3	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
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

Grade 5	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
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 Control Gain	ABCD Impact	Word-level interventions Control Gain	BCD Impact	Failure Free Reading Control Gain	A Impact	Spell Read Control Gain	B Impact	Wilson Reading Control Gain	C Impact	Corrective Reading Control Gain	D Impact
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 #
Aimsweb	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 Control Gain	ABCD Impact	Word-level interventions Control Gain	BCD Impact	Failure Free Reading Control Gain	A Impact	Spell Read Control Gain	B Impact	Wilson Reading Control Gain	C Impact	Corrective Reading Control Gain	D Impact
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 * #
Aimsweb	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.

xvii

Table 4

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

Grade 3	Baseline	All Interventions Control Gain	ABCD Impact	Word-level interventions Control Gain	BCD Impact	Failure Free Reading Control Gain	A Impact	Spell Read Control Gain	B Impact	Wilson Reading Control Gain	C Impact	Corrective Reading Control Gain	D Impact
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	Baseline	All Interventions Control Gain	ABCD Impact	Word-level interventions Control Gain	BCD Impact	Failure Free Reading Control Gain	A Impact	Spell Read Control Gain	B Impact	Wilson Reading Control Gain	C Impact	Corrective Reading Control Gain	D Impact
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

3rd Grade	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		
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	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		
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-Average for Control Group at follow-up).

Note: Gap defined as (100-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

Third Grade Gains in Word Attack

ᵃTreatment group gain
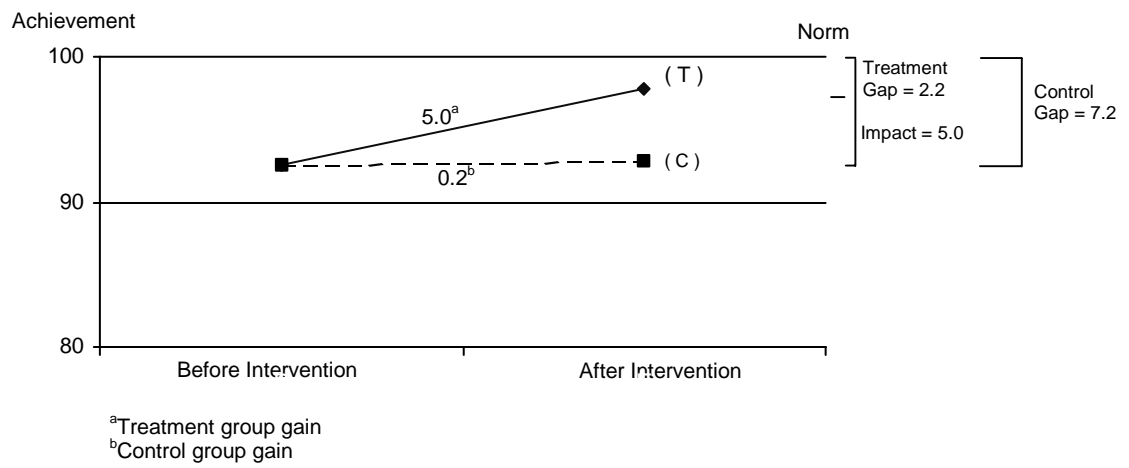ᵇControl group gain

Figure 2

Third Grade Gains in Phonemic Decoding Efficiency
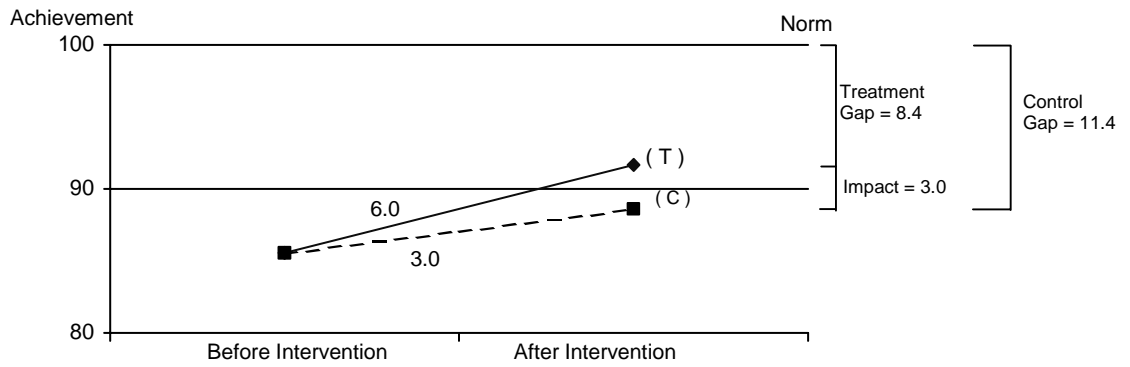

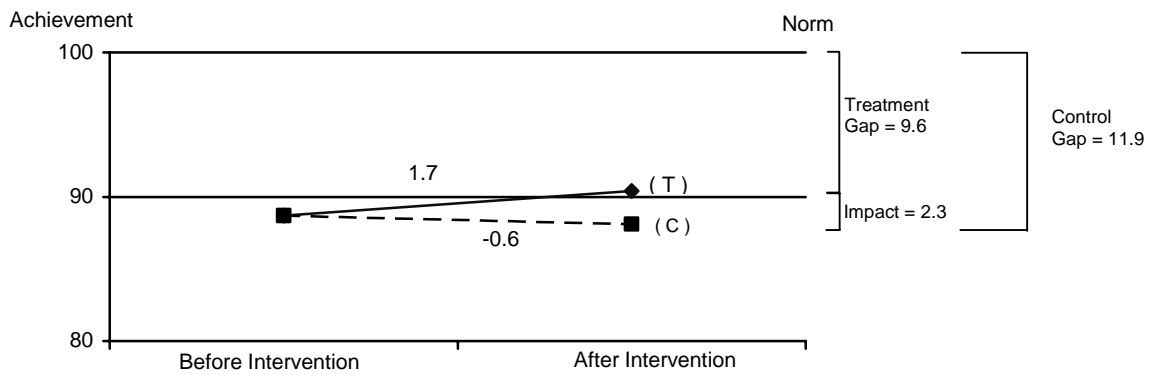Figure 3

Third Grade Gains in Word Identification

Figure 4

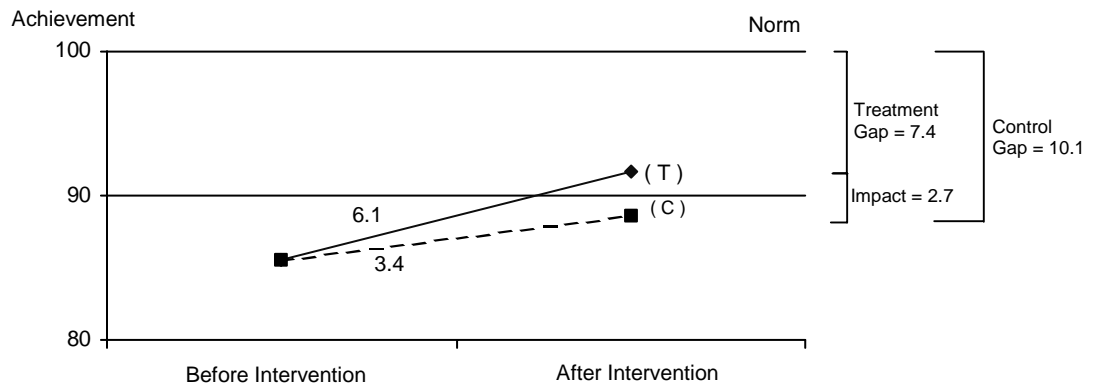Third Grade Gains in Sight Word Efficiency

Figure 5

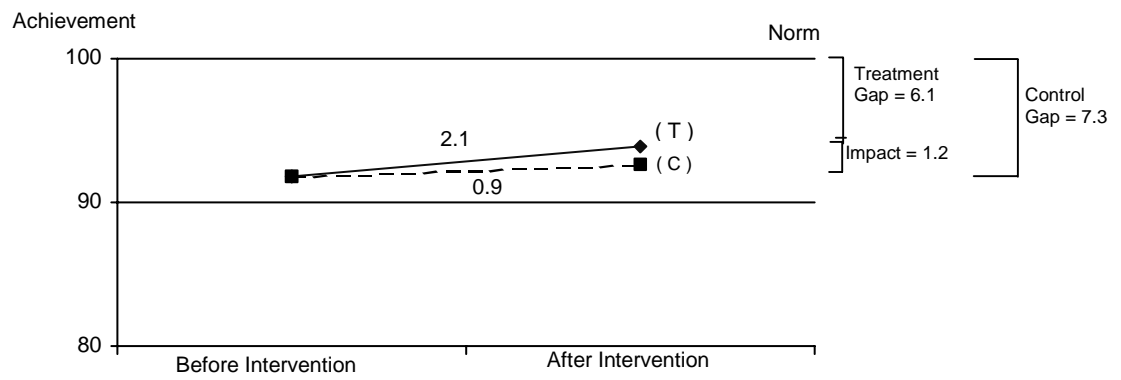Third Grade Gains in Passage Comprehension
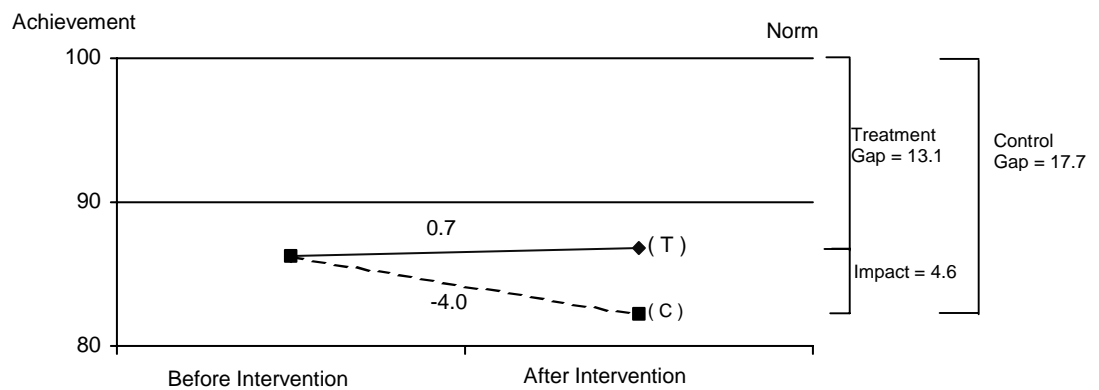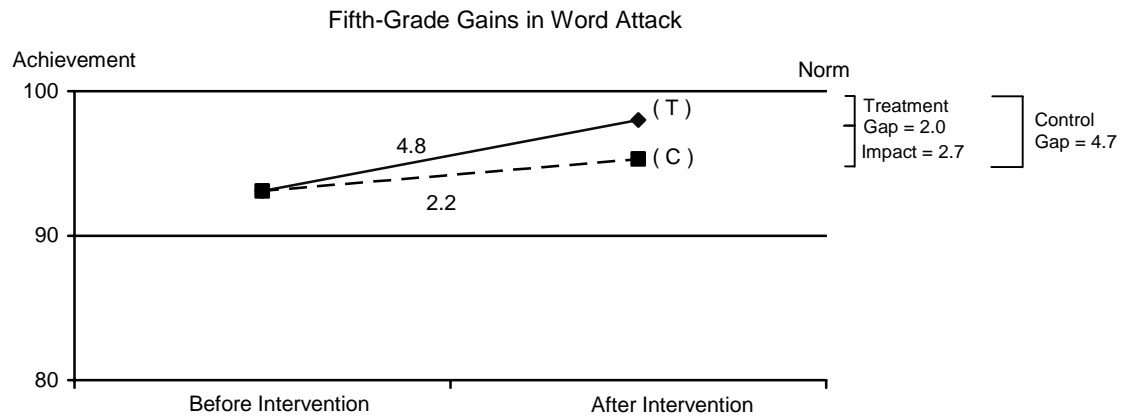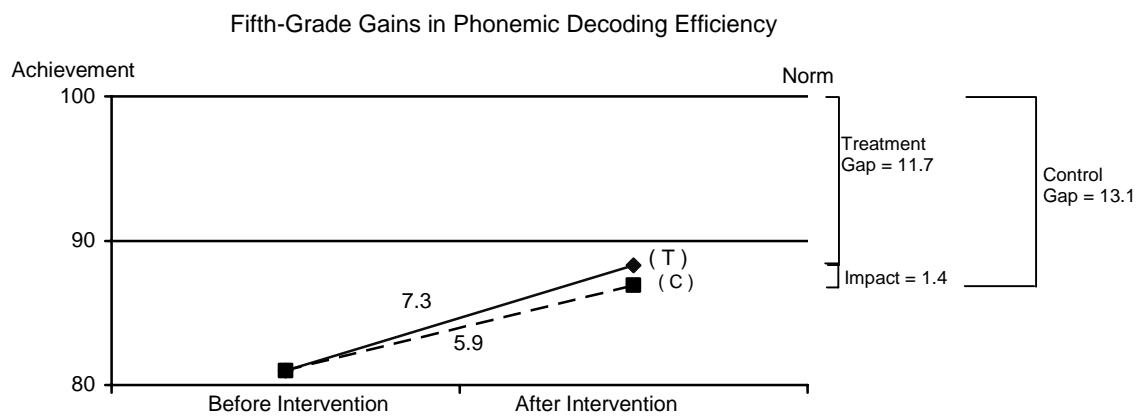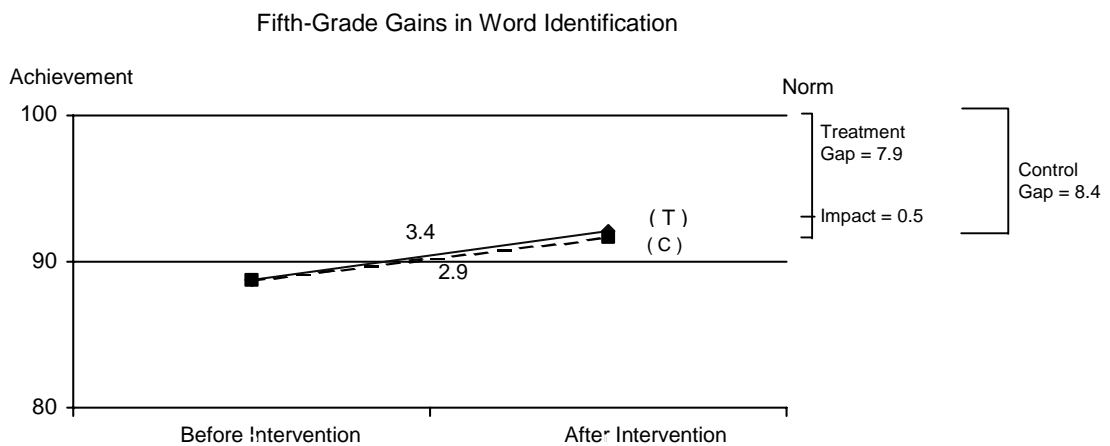

Figure 6

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
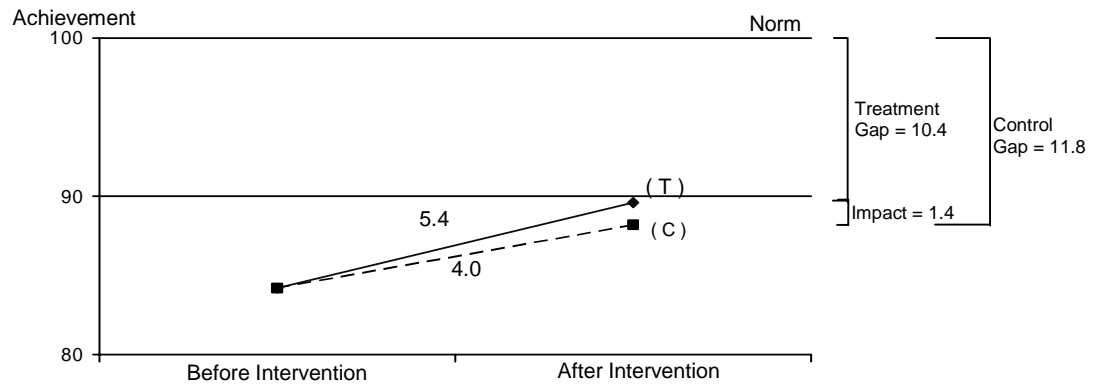
Fifth-Grade Gains in Sight Word Efficiency
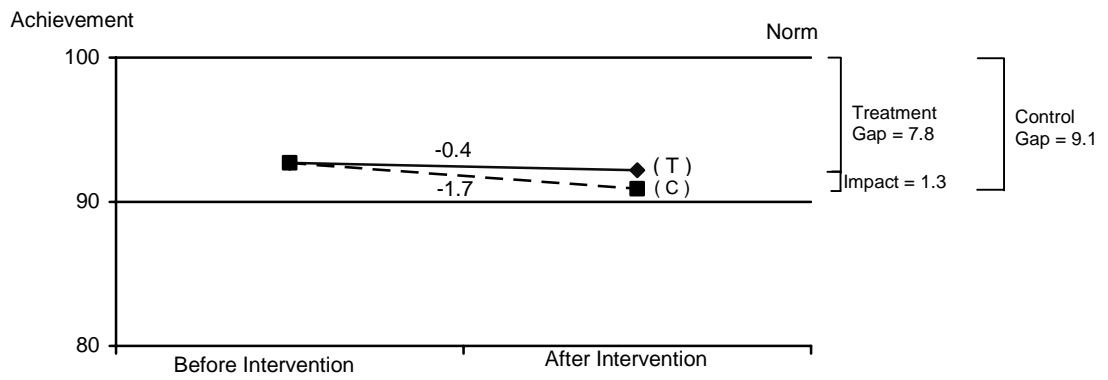

Figure 11

Fifth-Grade Gains in Passage Comprehension

Figure 12

Fifth-Grade Gains in GRADE Test