National Assessment of Title I
Final Report

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

A Report Prepared for IES
by the Corporation for the Advancement of Policy Evaluation

Under IES Grant R305U030002A

OCTOBER 2007

ACKNOWLEDGMENTS

This report reflects the contributions of many institutions and individuals. We would like to first thank the study funders. The Institute of Education Sciences of the U.S. Department of Education and the

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The research team for this evaluation consists of two grantees: Florida State University with Joe Torgesen as Principal Investigator, and the Corporation for the Advancement of Policy Evaluation with David Myers and Allen Schirm as Principal Investigators. None of these organizations or their key staff has financial interests that could be affected by findings from the evaluation of the four reading interventions reported here. No one from the Scientific Review Board, convened to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

EXECUTIVE SUMMARY

KEY EVALUATION QUESTIONS AND STUDY DESCRIPTION

Evaluation Context

According to the National Assessment of Educational Progress (U.S. Department of Education 2006), 36 percent of fourth graders read below the basic level. Such literacy problems can worsen as students advance through school and are exposed to progressively more complex concepts and courses. While schools often are 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.

were third and fifth graders. In this second (and last) report from the evaluation, we present estimates of impacts on scores from the same tests as of the end of the following year, when most of the students were fourth and sixth graders. We also present estimates of impacts on PSSA scores. However, the scores that we were able to obtain were from the tests taken the previous year, that is, bing the intervention year when the students were third and fifth graders. In other words, this report presents impacts at two points in time: late March to early April of the intervention year (for PSSA scores) and the end of the following school year (for the tests that we administered). Because most of the third and fifth graders in the first year are, respectively, fourth and sixth graders in the following year, we refer to and analyze student outcomes according to their "grade cohorts," which refer to the students' grade level when they entered the evaluation. Specifically, we present estimates for the "third-grade cohort" and the "fifth-grade cohort."

We provide detailed estimates of 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. An impact of the four interventions combined is not the impact of implementing the four interventions simultaneously. Rather, it can be interpreted as the impact of providing a struggling reader in third or fifth grade with the opportunity to receive a substantial amount of instruction in small groups with reasonably well-trained teachers, although as noted elsewhere, the content and instructional focus across the four interventions varied considerably. Such an impact is of greatest relevance to federal and state policymakers who can support broad programmatic approaches to instruction but cannot generally

In practice, impacts were estimated using hierarchical linear models, with separate models for the third- and fifth-grade cohorts. The models include a student-level model and a school-level model. In the student-level model, we include an indicator for treatment status and 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 includes 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.

Our key findings are as follows:

- ***The interventions improved some reading skills.*** For students in the third-grade cohort, the four interventions combined had impacts on phonemic decoding, word reading accuracy and fluency, and reading comprehension, although impacts were not detected for all measures of accuracy and fluency or comprehension (see Table 2). For students in the fifth-grade cohort, the four interventions combined improved phonemic decoding on one measure, but led to a small reduction in oral reading fluency. The three word-level interventions combined had similar impacts to those for all four interventions combined, although they did not have an impact on either measure of comprehension for students in the third-grade cohort, and they did have impacts on both measures of phonemic decoding for students in the fifth-grade cohort. For students in the third-grade cohort, Failure Free Reading (the only word level plus comprehension program) had impacts on one measure of phonemic decoding, two of the three measures of word reading accuracy and fluency, and one measure of comprehension. However, this intervention did not have any impacts for students in the fifth-grade cohort.

- ***The interventions did not improve PSOeD[fosion 0)5n(s 4..25(did not h)5vTC2_0 1 Tf-006 Tc 0Tw 11.3 0***

The key findings presented in this report for the seven tests administered for this study one year after the interventions ended are similar to the findings from the end of the intervention year. In our earlier report (Torgesen et al. 2006) we found that the four interventions combined and the three word-level interventions had impacts for students in the third-grade cohort on phonemic decoding, word reading accuracy and fluency, and reading comprehension. We found fewer significant impacts for students in the fifth-grade cohort than for students in third-grade cohort. Also, for the four interventions combined, the reading gaps for students in the intervention group were generally smaller than the gaps for students in the control group.

Grade 3 Cohort	Baseline	Control Gain		Control Gain		Control Gain		Control Gain		Control Gain		Control Gain	
Word Attack	92.4	-0.3	5.3 *	0.3	5.5 *	-2.2	4.9 *	-0.3	5.4 *	0.5	5.8 *	0.7	5.2 *
TOWRE PDE	85.5	1.1	4.0 *	1.0	4.9 *	1.3	1.3						

Grade 3 Cohort						
PSSA Reading	-15.6	-3.8	-51.1 *	-39.9	52.5	-23.8
PSSA Math	20.2	14.2	38.4	-15.5	56.6 *	1.4
Sample Size	329	240	89	92	71	77
Grade 5 Cohort						
PSSA Reading	-27.3 *	-25.3	-33.4 *	-30.0	-23.8	-22.1
PSSA Math	-28.8 * #	-34.0 * #	-13.4	-20.1	-56.4 * #	-25.4 *
Sample Size	408	280	128	102	92	86

Note: According to the first row of estimates, students in the third-grade cohort assigned to the Failure Free Reading intervention achieved a standardized score on the PSSA Reading test that was 51.1 points lower than the average score achieved by the students in the control group, a sts lower tthe the therag score ach2(o)-1(rm)1udentts lowigned to ttts 4r(icall)7(i(corets)1)0 Td(8tts),03 1ce, as i(w)26(as 51.1)tistcore ach2(o)-(t)7(ic)7(all)7(y)20

* Impact is statistically significant at the 0.05 level.
# Impact is statistically different form the 3rd grade impact at the 0.05 level.

Figure 7

Gap Reduction for Fifth-Grade Cohort: Word Attack

Percent Reduction in gap =
(2.7/5.3)*100 = 50%

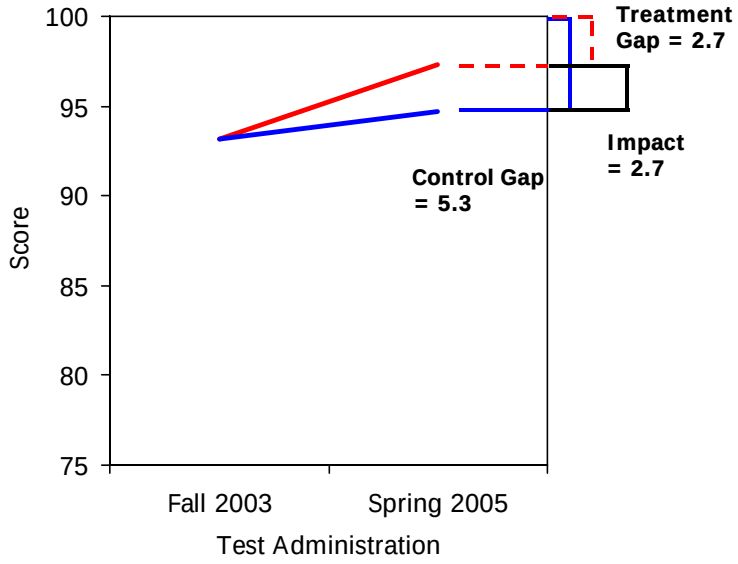


Figure 9

Gap Reduction for Fifth-Grade Cohort: Word Identification

Percent Reduction in gap =
(-0.6/8.3)*100 = -7%

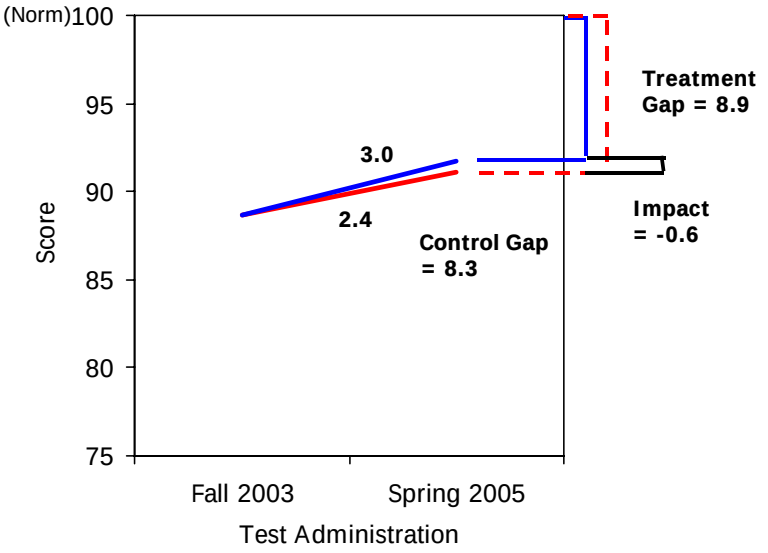
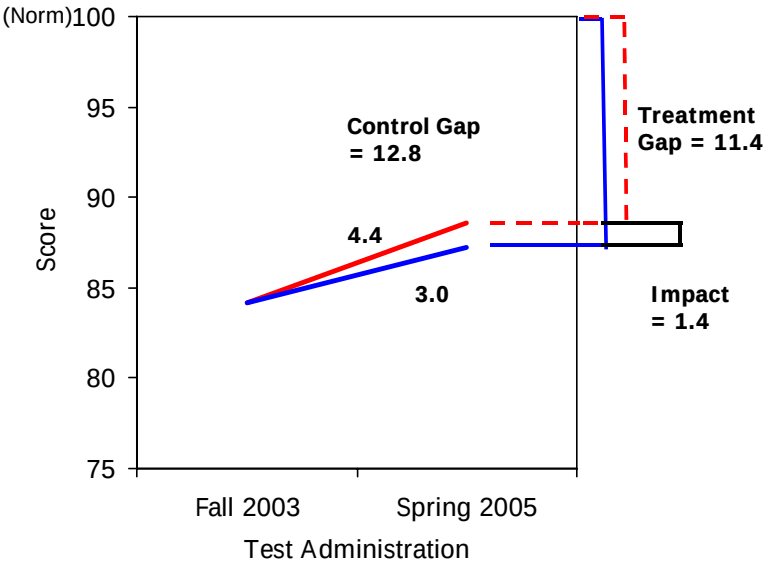
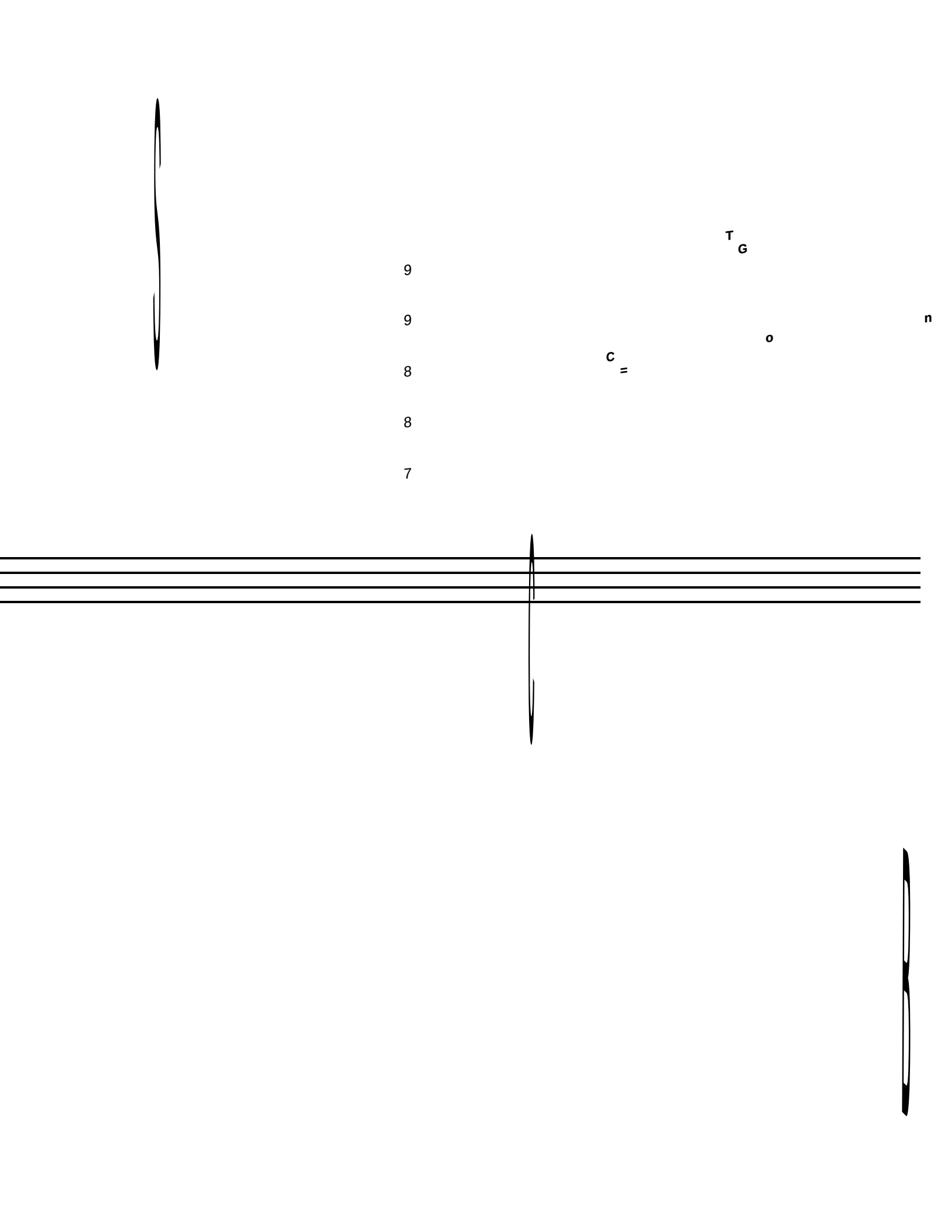


Figure 10

Gap Reduction for Fifth-Grade Cohort: TOWRE SWE

Percent Reduction in gap =
(1.4/12.8)*100 = 11%




T
G
9
9
n
o
C
=
8
8
7

I. INTRODUCTION

A. OVERVIEW

According to the National Assessment of Educational Progress (U.S. Department of Education 2006), 36 percent of fourth graders read below the basic level. Unfortunately, such 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, and the consequences 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.

To address this problem, many school districts have created remedial programs that aim to produce, on

II. DESIGN AND IMPLEMENTATION OF STUDY

This evaluation has two main elements: (1) an impact study, and (2) an implementation study. The implementation study has examined the instruction provided by the four interventions and the instruction provided outside of the interventions to both the students who participated in the interventions and those who did not. We describe the design and main findings of that study in Torgesen et al. (2006). We summarize the findings and present some additional findings in the next chapter.

This chapter focuses on the impact study. The impact study is based on a scientifically rigorous design—an experimental design that uses random assignment at two levels: (1) schools were randomly assigned to one of the four interventions; and (2) within each school, eligible children in grades three and five were randomly assigned to a treatment group or to a control group. Randomization at the school-level was done so that the interventions would be implemented within similar schools. Randomization at the student-level ensures that the students in the

within four strata defined by the percentage of students eligible for free or reduced-price school lunch.[5] One school unit (consisting of two schools) dropped out of the study after randomization, but before it learned of its random assignment, leaving 31 school units and 50 schools in the study.[6,7]

To assess the similarity of the intervention groups after randomly assigning schools, Table II.1 shows the distribution of school unit-level covariates across the four groups of school units assigned to each intervention. Torgesen et al. (2006) also compared the schools in the study with other schools in the AIU and with schools nationwide. Tables II.2 and II.3 present comparisons based on student-level covariates, and the final columns of each of those tables also show tests of significance for differences in student-level covariates across the four interventions. The only two significant differences in the school unit-level covariates across the four interventions are both attributable to differences in school size. By chance, five of the six smallest schools were assigned to Wilson Reading, so some of the variables directly related to enrollment (total enrollment and average class size) differ across the four interventions. On student-level covariates, we observe a few differences. With just 32 school units randomized, however, it is not surprising to observe such differences among the four groups.[8] While small differences could affect the inferences we draw from the impact analysis when comparing interventions, our impact analyses are based on the differences in reading achievement for students in treatment and control groups within school units, rather than between school units. Thus, small o some of thtie draw fm

- ***Identify Potentially Eligible Students***. Teachers in the 50 schools identified 1,576 struggling readers in third or fifth grade for screening. Nearly all (1,502) of these students were screened.[10]

School Characteristics	Failure Free Reading	Spell Read	Wilson Reading	Corrective Reading	
Measurements of School Size					
Total enrollment	506	563	389	508	*
Average enrollment per grade	118	113	68	118	
Number of grades in school	5	5	6	5	
Both 3rd and 5th grades in school	0.88	0.63	1.00	0.63	
Number of 3rd grade classes	4.4	5.0	3.4	4.4	
Number enrolled in 3rd grade	110	118	69	95	
Number of 5th grade classes	5.9	4.6	3.2	5.7	
Number enrolled in 5th grade	153	116	69	144	
Average class size	25	24	21	23	*
Characteristics of Students in the School					
Fraction eligible for free or reduced price lunch	0.35	0.36	0.40	0.34	
Fraction of students who leave during the year	0.04	0.03	0.09	0.09	
Fraction white	0.85	0.70	0.76	0.82	
Fraction African American	0.14	0.29	0.23	0.16	
0. 5	0.10	0.76	0.		

Baseline Means	Treat.	Cont.		Treat.	Cont.		Treat.	Cont.		Treat.	Cont.			
Student Characteristics														
Age	10.6	10.6		10.9	10.7		10.8	10.5	*	10.7	10.6			
Male (%)	53	51		55	58		54	66		49	59			
Hispanic (%)	a	a		a	a		a	a		a	a			
Race--White (%)	78	83		76	67		55	59		83	86			
Race--African American (%)	22	17		24	33		45	41		18	14			
Race--Other (%)	a	a		a	a		a	a		a	a			
Family income less than $30,000 (%)	41	50		51	54		73	47	*	32	46			
Family income between $30,000 and $60,000 (%)	43	33		39	34		23	36		43	36			
Family income over $60,000 (%)	16	17		10	12		a	a	*	25	18			
Eligible for free or reduced price lunch (%)	42	46		53	41	*	58	41	*	41	47			
Has any learning or other disability (%)	27	36		29	30		32	28		29	29			
Mother has bachelor's degree or higher (%)	12	17		5	9		a	a	*	15	29	*		
Screening Tests														
TOWRE Sight Word Efficiency	84.1	85.3		84.1	85.3		83.1	85.0		82.7	83.5			
TOWRE Phonemic Decoding Efficiency	81.7	79.8	*	78.1	79.6		82.3	82.3		80.1	81.2			#
Peabody Picture Vocabulary Test--Revised	94.8	95.1		92.3	92.6		91.8	99.7	*	95.2	98.3	*		#

Baseline Tests

TOWRE Shonem7c Decoding Efficiency

Table II.4

Baseline Characteristics of Full Sample and Sample for Three Word-level Interventions, by Treatment Status: 3rd Grade Analysis Sample

Baseline Means	All Interventions			Word-level Interventions		
	Treatment	Control		Treatment	Control	
Student Characteristics						
Age	8.7	8.7		8.7	8.7	
Male (%)	56	47		57	43	
Hispanic (%)	2	3		a	a	
Race--White (%)	68	72		65	69	
Race--African American (%)	32	28		35	31	
Race--Other (%)	a	a		a	a	
Family income less than $30,000 (%)	47	50		50	53	
Family income between $30,000 and $60,000 (%)	36	33		30	29	
Family income over $60,000 (%)	17	18		20	18	
Eligible for free or reduced price lunch (%)	43	47		42	46	
Has any learning or other disability (%)	34	33		33	29	
Mother has bachelor's degree or higher (%)	13	13		12	14	
Screening Tests						
TOWRE Sight Word Efficiency	85.6	83.3	*	85.9	83.7	*
TOWRE Phonemic Decoding Efficiency	85.4	85.7		85.9	85.9	
Peabody Picture Vocabulary Test--Revised	94.2	94.6		94.6	94.6	
Baseline Tests						
WRM Word Identification	89.7	87.7	*	89.9	87.8	*
TOWRE Phonemic Decoding Efficiency	86.0	85.0		86.6	85.3	*
WRM Word Attack	93.0	91.8		94.0	92.8	
TOWRE Sight Word Efficiency	88.2	84.9	*	88.5	85.1	*
AIMSweb (Raw score)	44.3	37.6	*	46.5	39.0	*
WRM Passage Comprehension	93.5	89.7	*	94.3	90.2	*
GRADE	88.0	84.3	*	88.6	84.0	*
Woodcock Johnson Spelling	89.9	87.2		89.7	87.4	
Woodcock Johnson Calculation	96.4	94.3	*	97.7	93.4	*
Sample Size	203	126		152	88	

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

Note: All test scores are shown as standard scores, unless otherwise indicated. All standard scores have mean 100 and standard deviation 15. The mean raw scores for AIMSweb tests, administered to students across the country in the fall in the school years 2000-2001, 2001-2002, and 2002-2003, were 75 and 112 for third and fifth graders, respectively. The respective standard deviations were 39 and 47.

* Difference between treatment and control groups is statistically significant at the 0.05 level.

a Values suppressed to protect student confidentiality.

Baseline Means	Treatment	Control		Treatment	Control	
Student Characteristics						
Age	10.7	10.6	*	10.8	10.6	*
Male (%)	53	58		53	61	
Hispanic (%)	a	a		a	a	
Race--White (%)	73	74		72	71	
Race--African American (%)	27	26		28	29	
Race--Other (%)	a	a		a	a	
Family income less than $30,000 (%)	48	49		51	49	
Family income between $30,000 and $60,000 (%)	38	35		36	35	
Family income over $60,000 (%)	14	16	*	14	16	*
Eligible for free or reduced price lunch (%)	48	44		51	43	
Has any learning or other disability (%)	29	31		30	29	
Mother has bachelor's degree or higher (%)	8	18	*	7	18	*
Screening Tests						
TOWRE Sight Word Efficiency	83.5	84.8		83.3	84.6	
TOWRE Phonemic Decoding Efficiency	80.5	80.7		80.1	81.0	
Peabody Picture Vocabulary Test--Revised	93.6	96.4	*	93.2	96.8	*
Baseline Tests						
WRM Word Identification	88.2	89.1		87.5	89.2	
TOWRE Phonemic Decoding Efficiency	80.7	81.3		80.2	81.2	
WRM Word Attack	92.7	93.7		92.4	94.0	
TOWRE Sight Word Efficiency	83.7	84.7		83.5	84.3	
AIMSweb (Raw score)	77.2	76.8		76.6	77.2	
WRM Passage Comprehension	91.5	93.7		91.2	94.3	*
GRADE	90.3	92.1		89.9	92.0	
Woodcock Johnson Spelling	90.8	91.1		89.8	90.7	
Woodcock Johnson Calculation	94.1	94.3		94.1	94.6	
Sample Size	224	176		162	110	

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

* Difference between treatment and control groups is statistically significant at the 0.05 level.

- ***Randomly Assign Eligible Students to the Treatment and Control Groups***

analysis.[18]

92 for Passage Comprehension. These baseline scores for word-level skills are much higher than corresponding scores from a set of 13 intervention samples recently reviewed by Torgesen (2005). The students in those studies were of approximately the same ages as those in the present study, and their average baseline standard score for Word Attack was 75 and their average baseline score for Word

- Oral Reading Fluency subtest from Edformation, Inc., (Howe and Shinn 2002). The text of this report refers to these 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; Williams 2001)

For all tests except the AIMSweb passages, the analysis used grade-normalized standard scores, which indicate where a student falls within the overall distribution of reading ability among students in the same grade.[21,22] 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.[23] For the AIMSweb passages, the score used in this analysis is the median correct words per minute from three grade-level passages. (See the note on Table II.2 for more information about the means and standard deviations for the scores on the AIMSweb tests.) Table II.8 shows estimates of test reliability, and Tables II.9 and II.10 present correlations between tests for the third-grade students and fifth-grade students, respectively. The shaded boxes in Tables II.9 and II.10 indicate tests that measure similar constructs: tests measuring phonemic

for the WRMT-R (Woodcock 1998) reports a correlation between the Word Identification measure and Passage Comprehension measure of 0.67 for third graders and 0.59 for fifth graders. The lack of a strong correlation between the two measures of reading comprehension may reflect several differences in the way the tests are administered and the types of required responses.

b. Measures of Spelling and Mathematics Calculation Ability Administered for the Evaluation

The spelling and calculation subtests from the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, and Mather 2001) assessed spelling and mathematics calculation abilities. Table II.8 includes estimated reliabilities for these tests as well as the seven reading tests.

c. Measures of Reading and Mathematics from the PSSA

Table II.8

Tests Administered for the Second Follow-Up (End of the 2004-05 School Year)

Test	Reliability
Measures of Reading	
Phonemic Decoding	
Woodcock Test-R (WRMT-R) Word Attack (WA)	0.90[a]
Test of Word Reading Efficiency (TOWRE) Phonemic Decoding Efficiency (PDE)	0.93[b]
Word Reading Accuracy and Fluency	
WRMT-R Word Identification (WI)	0.94[a]
TOWRE Sight Word Efficiency (SWE)	0.95[b]
Aimsweb Oral Reading Passages (AIMS)	0.92[b]
Reading Comprehension	
WRMT-R Passage Comprehension (PCG)	0.82[a]
Group Reading Assessment and Diagnostic Evaluation Passage Comprehension (GRADE)	Grade 3: 0.88[c] Grade 5: 0.90[c]
Other Tests	
Woodcock Johnson Tests of Achievement-III (WJ-III)	
Spelling	0.89[c]
Calculation	0.85[c]

[a] Split-half reliability
[b] Alternate-form reliability
[c] Internal consistency reliability

b. School Records Form

At the end of the 2004-05 academic year, we obtained data on each student using a school records form. We collected information on enrollment, attendance, and suspensions; characteristics such as limited English proficiency status, eligibility for the free or reduced-price lunch program, and disabilities; reading services; Individual Education Plan (IEP) or Service Agreement specifications; grade promotion and retention; course grades; and reading and math standardized test scores.

EXHIBIT 1. SECOND FOLLOW-UP STUDENT PERFORMANCE MEASURES

READING MEASURES

Phonemic Decoding

- **Word Attack** subtest from the Woodcock Reading Mastery Test-Revised (WRMT-R; Woodcock 1998) requires students to pronounce printed nonwords that are spelled according to conventional English spelling patterns.

- **Phonemic Decoding Efficiency** subtest from the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, and Rashotte 1999) requires students to pronounce nonwords from a list of increasing difficulty as fast as they can. The score is the number of words correctly pronounced within 45 seconds.

Word Reading Accuracy and Fluency

- **Word Identification** subtest from the WRMT-R requires students to pronounce real words from a list of increasing difficulty. The child's score is the total number of words read correctly before reaching a ceiling, which is determined when the child makes a specific number of errors in a row.

- **Sight Word Efficiency** subtest from the TOWRE requires students to pronounce real words from a list of increasing difficulty as fast as they can. The score is the number of words correctly pronounced within 45 seconds.

- **Oral Reading Fluency** subtest from Edformation, Inc., (Howe and Shinn, 2002) requires students to read three passages at their grade level (third or

III. IMPLEMENTATION ANALYSIS

The purpose of this evaluation is to estimate the impact of four reading interventions when they are delivered with as much fidelity and skill as can be attained in a standard school setting. Our procedures to ensure high quality implementation of the interventions included careful selection of teachers to deliver the interventions, training and supervision of intervention teachers by the program developers, and the use of a full-time study coordinator, whose duties included working with school personnel to facilitate the scheduling of intervention sessions and minimize disruptions so that each student could receive at least 100 hours of instruction. We then collected information to evaluate the quality and fidelity of the intervention implementations, as well as to understand how the interventions fit into the overall reading instruction for each child. A detailed discussion of our findings from this assessment, and a description of the procedures for selecting, training, and supporting the intervention teachers can be found in Torgesen et al. (2006). In this report, we summarize the key implementation findings from the prior report and present some new findings pertaining to students' hours of reading instruction in the year after the interventions.

As described in Torgesen et al. (2006), we used a variety of instruments to evaluate implementation. These included documentation of the amount of training received by each of the intervention teachers, daily attendance logs for all intervention sessions, video-tapes of a sample of intervention sessions, ratings of the intervention teachers by program trainers and study coordinators, and questionnaires completed for each participating student by his or her classroom teacher. The video tapes were used in three separate analyses: an analysis of program fidelity and general teacher quality, a verification of session length (which had implications for total hours of instruction), and an analysis of intervention program content. The classroom teacher surveys were used to describe each student's total reading instruction; these surveys were completed for both the intervention year and the year following the intervention.

Later in this section, we present an integrated discussion of our findings regarding hours of total reading instruction during both years of the study. We begin, however, with a brief summary of our other key findings from the first year of the study and a description of the instructional elements and procedures for each of the four interventions.

A. SUMMARY OF KEY IMPLEMENTATION FINDINGS

Hours of instruction. The large majority (93 percent) of students in the treatment group received at least 80 hours of intervention instruction. This represents a sustained and substantial level of exposure to intensive instruction, even though only 14 percent of intervention students received the intended dose of 100 hours. There were no significant differences in average hours of instruction across interventions, although fifth-grade students received fewer hours of intervention (88 hours) on average than did third-grade students (93 hours).

Heterogeneity of instructional groups. Due to the practical constraints imposed by the incidence and diversity of reading difficulties among third and fifth graders in the AIU schools participating in this evaluation, the instructional groups formed for the intervention were heterogeneous with regard to their beginning word-level skills. At each grade level, the average difference between the highest and lowest baseline Word Attack scores among the three students in an instructional group was about one standard deviation. Nonetheless, the program developers indicated in follow-up conversations that this amount of

within-group heterogeneity was not unusual in comparison with what they normally observe when delivering their interventions in other settings.[24]

Training of intervention teachers. Representatives of the four reading programs trained the intervention teachers. On average, teachers received almost 70 hours of professional development over the course of the intervention, starting with five days of intensive training for all teachers in August 2003. The total amount of professional development varied across the reading programs, but all of the program providers agreed that the amount of training and professional development equaled or exceeded what they would typically deliver to new teachers in a school setting.

Trainer ratings of intervention teachers. The trainers from each reading program rated the teaching performance of teachers under their supervision. According to the trainers, the average instruction teacher's performance fell somewhere in the top half among similarly experienced teachers whom they had observed. In addition, the trainers' average ratings on five dimensions of program fidelity and three dimensions of general teacher quality were well above the satisfactory level for all dimensions and all programs.

Video analysis of intervention teachers. Each teacher was videotaped twice over the course of the intervention, and the videos were analyzed for adherence to program guidelines. Each program had slightly different dimensions along which fidelity was assessed, as well as different criteria for judging adequacy. Deviations from criterion were judged by members of the evaluation team as minor, moderate, or extreme. Overall, there were no extreme deviations and relatively few moderate deviations. The moderate deviations that did occur were primarily with regard to time in session (most sessions ran shorter than anticipated) and fine points of program technique.

Time by activity analysis. The videotaped instructional sessions were also analyzed to determine how teachers allocated time across instructional activities. The analysis showed that the distribution of time between word-level and vocabulary/comprehension activities did not conform to the categorization of the interventions in the original study design (which was based on the description of instructional activities from the program providers). As a consequence, the programs were regrouped for analysis, with the three programs that devoted most of their instructional time to improving word-level reading skills grouped together.

More detailed information pertaining to these and other results from our implementation analysis can be found in Torgesen et al. (2006).

B. DESCRIPTION OF THE INTERVENTIONS

A description of the essential instructional elements and procedures of each of the four instructional methods, as they were implemented in this study, is provided below, along with results from Torgesen et al. (2006) about the relative amount of time devoted to instruction in word-level skills versus vocabulary and comprehension.

[24] Furthermore, we previously found no consistent patterns in the relationship between instructional group heterogeneity and students' reading outcomes (Torgesen et al. 2006).

pace based on initial placement testing and frequent criterion testing. Two levels of story books are available.
Students who show mastery at the second level progress to a related program called Verbal Master, which uses the same instructional principles but emphasizes vocabulary building and writing activities rather than passage reading. Verbal Master activities include listening to definitions and applications of target vocabulary words and interpreting and constructing sentences containing the target words. The curriculum also provides reinforcement exercises such as sentence completion and fill-in-the-blank activities as well as basic instruction in composition. Most of the third-grade students assigned to the Failure Free condition spent all of their instructional time working within the first and second level of story sequences. On the other hand, 65 percent of the fifth-grade students spent half or more of their instructional time in Verbal Master.

Estimated allocation of instructional time: 48 percent on word-level skills and 52 percent on comprehension/vocabulary.

C. TOTAL HOURS OF READING INSTRUCTION

In addition to the hours spent in the experimental treatment, students in the intervention group also received some reading instruction in their regular classrooms. Control students, in accordance with the study design, received the mix of reading services that would normally be provided by their schools. To better understand the treatment contrast, we examined total reading instruction provided to treatment and control students during the intervention year, and again during the following year. The intervention year findings were initially presented in Torgesen et al. (2006). They are repeated here, disaggregated by grade cohort, to facilitate comparison with results for the following year.

Hours of reading instruction were based on annual surveys, which were completed for each student in the study by his or her regular classroom teacher. The surveys included questions on the duration of reading instruction provided for that student during a typical week, the sizes of the groups in which instruction was delivered, and the types of professionals who provided the instruction. For students 5(he e)]int6npsuandJ

Figure III. 2

Average Hours of Reading Instruction per Week During Intervention Year: Fifth-Grade Cohort.

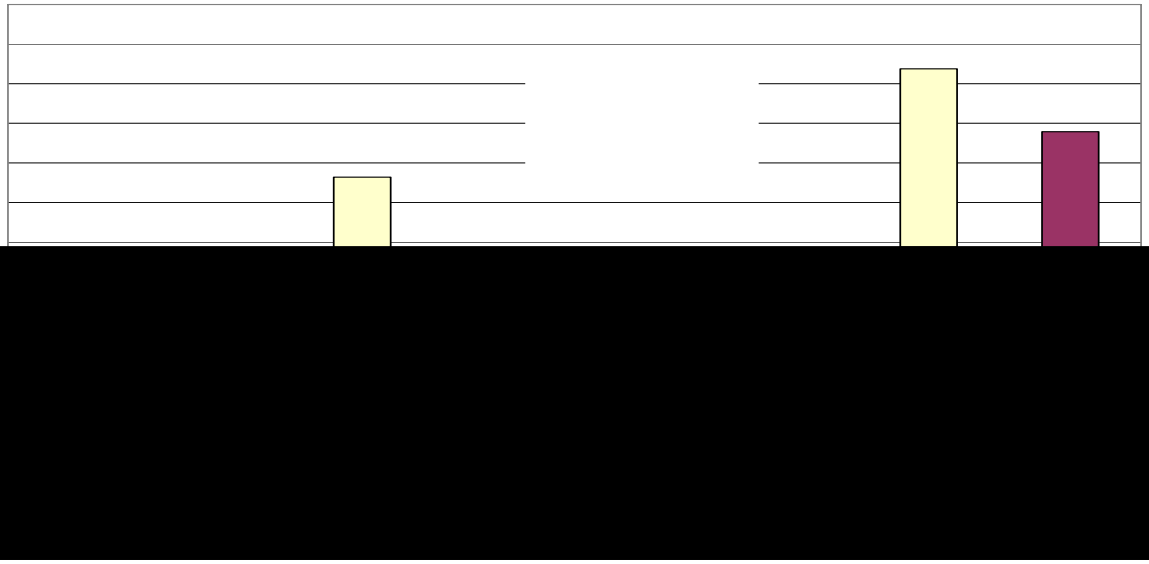
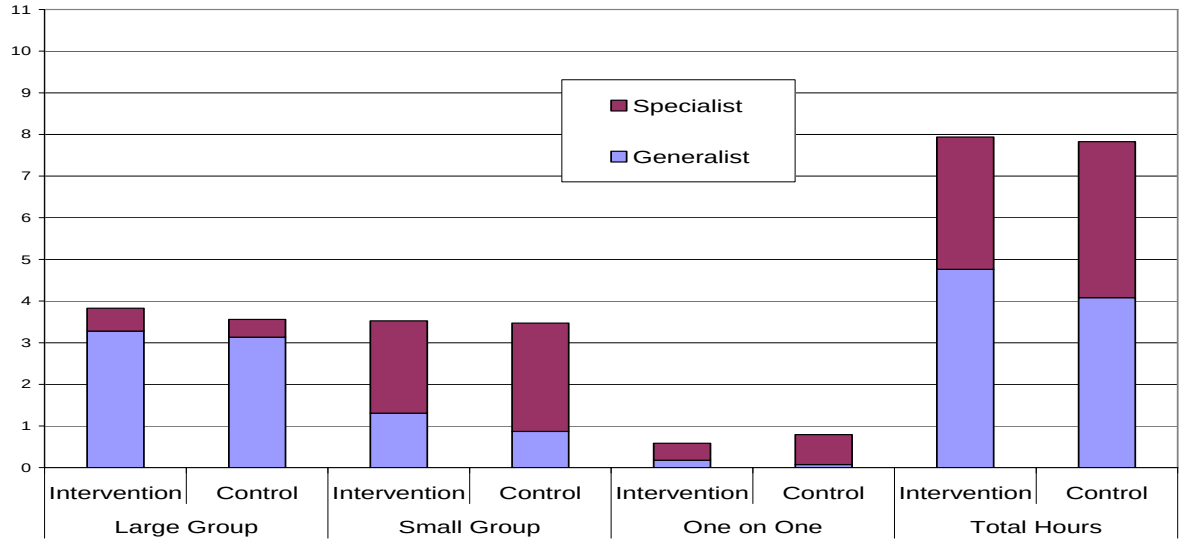


Figure III.3

Average Hours of Reading Instruction per Week During Year Following Intervention: Third-Grade Cohort.

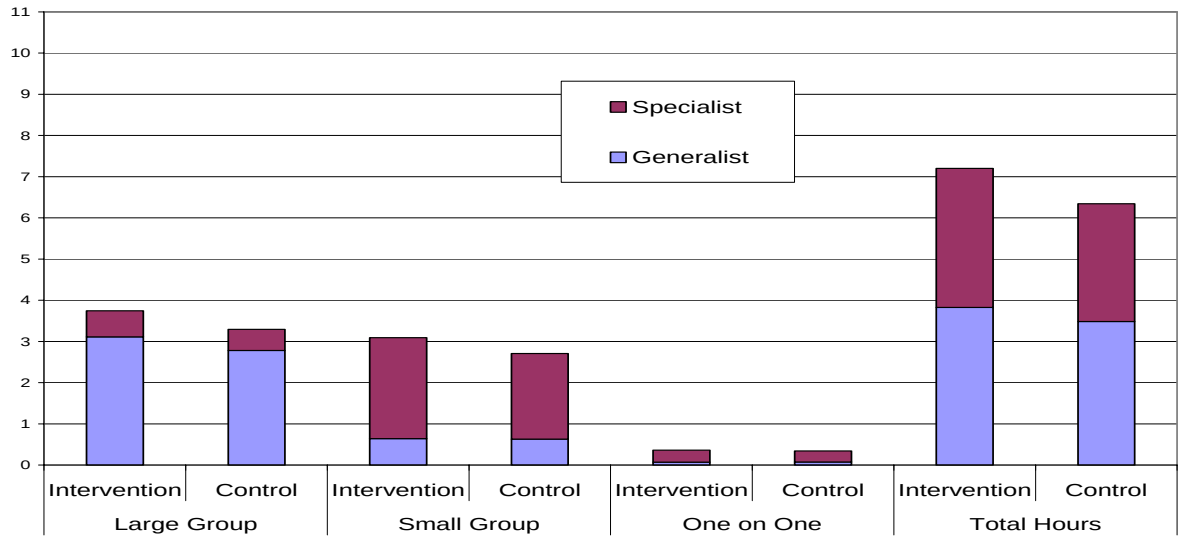


Note: A "Specialist" is defined as a teacher who is not a general education teacher. For example, specialists include special education and Title I teachers and reading specialists.

For the third-grade cohort (Figure III.3), there were no significant differences in the total hours of instruction received by the intervention and control groups during the year following the intervention (7.9 and 7.8 hours, respectively). Neither were there any significant differences by group size or by type of instructor.

Figure III.4

Average Hours of Reading Instruction per Week During Year Following Intervention: Fifth-Grade Cohort.



Note: A “Specialist” is defined as a teacher who is not a ge

IV. IMPACT ANALYSIS

The main objective of this evaluation is to estimate the impacts of the four interventions on students' reading skills. Specifically, we estimate the impacts of the four interventions combined, the three word-level interventions combined, and each of the four individual interventions for not only all students in the third-grade cohort and all students in the fifth-grade cohort, but also several key subgroups of students. In this chapter, we present the findings of our impact analysis after describing our estimation methods and technical and contextual issues pertaining to the interpretation of the impact estimates.

A. ESTIMATION METHOD

The experimental design can be described as a randomized blocks design with random assignment carried out at two levels. First, as discussed in Chapter II, we randomly assigned 32 school units to the four interventions within blocking strata determined by the percentage of students eligible for free or reduced-price school lunch.[25] Next, within schools, we randomly assigned eligible students within grade levels (third or fifth) to the treatment or control group. The resultant data have a hierarchical structure of students nested within school units.

To reflect the fact that students within a school unit are not independent, we estimated intervention impacts and standard errors using a weighted two-level hierarchical linear model (HLM) that allows for nested data.[26] The first level corresponds to students within school units and the second to the school units, accounting for the clustering (nonindependence) of students in school units.

Research has shown that the impacts of interventions may vary by age, and that older students experience more difficulty in improving their reading skills (Torgesen 2005). Therefore, we estimated impacts separately for the third- and fifth-grade cohorts (but test for differences between impacts for the two cohorts). The model is:

Level One: Student *i* within school unit *j*

$$y_{1ij} = \beta_{0j} + \beta_{1j} T_{ij} + \beta_{2j} y^{*}_{oij} + r_{ij} \qquad \text{(IV.1)}$$

[25] The sample includes 31 school units with about 730 students; one school unit dropped out of the study after random assignment, but before learning about the intervention to which it had been assigned.

[26] We also investigated a three-level model that includes a level for the clustering of intervention students in instructional groups. The results are similar to those obtained when using a two-level model; see Appendix F for details of the three-level model and the results. In most cases, st

Level Two: School unit (*j*)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}A_j + \gamma_{02}B_j + \gamma_{03}C_j + \sum_{l=1}^{3}\xi_{0l}P_{lj} + \mu_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}A_j + \gamma_{12}B_j + \gamma_{13}C_j + \sum_{l=1}^{3}\xi_{1l}P_{lj} + \mu_{1j} \qquad \text{(IV.2)}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}A_j + \gamma_{22}B_j + \gamma_{23}C_j + \sum_{l=1}^{3}\xi_{2l}P_{lj} + \mu_{2j}$$

where

$T_{ij} = 1$ if student i in school unit j was randomly assigned to the treatment group (intervention), and $T_{ij} = 0$ if student i in school unit j is in the control group;

$A_j = 1$ if school unit j was randomly assigned to the Failure Free Reading intervention, and $A_j = 0$ otherwise;

$B_j = 1$ if school unit j was randomly assigned to the Spell Read P.A.T. intervention, and $B_j = 0$ otherwise;

$C_j = 1$ if school unit j was randomly assigned to the Wilson Language Training intervention, and $C_j = 0$ otherwise;

$P_{1j} = 1$ if school unit j is in blocking stratum 1, and $P_{1j} = 0$ otherwise;

$P_{2j} = 1$ if school unit j is in blocking stratum 2, and $P_{2j} = 0$ otherwise;

$P_{3j} = 1$ if school unit j is in blocking stratum 3, and $P_{3j} = 0$ otherwise;

y_{1ij} = post-test score;

y^{*}_{0ij} = centered pretest score.

For our analyses, we use a centered pretest score:

$$y^{*}_{0ij} = y_{0ij} - \bar{y}_{..}, \qquad \text{(IV.3)}$$

where $\bar{y}_{..}$ is the weighted mean of the pretest score across all students in a given grade cohort in the evaluation sample. By mean-centering the pretest score (that is, the baseline score), we can interpret parameters and combinations of parameters in the level-one model as means for students with the average baseline test score. For example, the impacts, estimated as described below, are interpreted as the impact for a student in a given grade cohort (third or fifth) with a baseline test score equal to the average baseline test score across students in that grade cohort.

The level-one model (Equation (IV.1)) relates a student's post-intervention test score to a treatment indicator, the student's pretest score, and a residual term (unexplained variation). The level-two model (Equation (IV.2)) relates the level-one parameters (coefficients $\beta_{0j}, \beta_{1j},$ and β_{2j}) to indicators for the interventions to which the school units were randomly assigned and the blocking strata. The interventions Failure Free Reading, Spell Read, Wilson Reading, and Corrective Reading are denoted as A, B, C, and D, respectively.[27] The blocking strata grouped school units into four approximately equal-sized groups based on the percentage of students eligible for free or reduced-price school lunch (FRPL). We represent the four blocking strata with three dummy variables, where each dummy variable equals 1 for school units that belong to that blocking stratum, and zero otherwise.[28]

The main parameters of interest in our two-level model are those from which we estimate the impacts of the interventions on students' reading skills, where an impact is defined as the regression-adjusted difference in the average achievement scores for the treatment and control groups.[29,30] Such an impact shows how much difference an intervention will make if it is made available to students with characteristics similar to those of the students in the evaluation sample. This is the most robust estimate of program impact because it involves the fewest assumptions when estimating the impact. By imposing more assumptions, which might not be valid, we could also estimate the intervention impacts on those who participated in the interventions, and on those who participated substantially, receiving at least 80 hours of instruction, for example. Given that almost all students in the treatment group received some of the treatment, and that a very large percentage received 80 or more hours of instruction, the results are similar, regardless of the definition of impacts, as discussed in more detail below.

From the HLM model, we estimate impacts for each of the four interventions.[31] We also estimate the impact of assignment to any of the interventions—denoted as the combined intervention impact (ABCD)—as the average of the four intervention impacts.

As explained earlier in this report, we had originally intended to group the four intervention programs into two intervention classes: word-level interventions and word-level plus comprehension/vocabulary interventions. However, the time-by-activity analysis indicated that such a categorization was not accurate. In actuality, three of the interventions, Corrective Reading, Spell Read, and Wilson Reading, were appropriately grouped as phonemically oriented word-level interventions, while the fourth, Failure Free Reading, provided non-phonemically oriented support for reading accuracy and fluency, along with instruction in comprehension and vocabulary. For the analyses reported here, we consider impacts for:

---

[27] The listed order of the interventions and labels A, B, C, and D are arbitrary and not related to the performance of the interventions. In the hierarchical model, we can represent the four interventions with three dummy variables: A, B, and C. Intervention D is represented when the dummy variables for interventions A, B, and C all equal zero (i.e., A=B=C=0).

[28] When estimating impacts, we weight the blocking strata effects equally.

[29] Our analyses compare the treatment students in each intervention to control students in the same schools, which require minimal assumptions about how the controls differ across interventions, compared with an analysis that pools all of the controls. The impacts refer to the average impacts across school units and to students with the average baseline test score.

[30] Appendix D provides details on deriving the impacts from estimated model parameters.

[31]

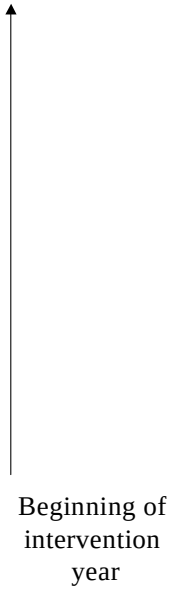
1. All interventions combined (ABCD)
2. The three word-level interventions combined: Spell Read, Wilson Reading, and Corrective

in which the interventions were operating: (1) where the students began in terms of reading ability just prior to the interventions, (2) how much improvement the students would have made in the absence of the interventions, and (3) the amount of the interventions that treatment and control students actually received.

We illustrate the first two elements using a hypothetical example, in Figure IV.1.[35] At the beginning of the intervention year, all students in the intervention (represented by "T") and control (represented by "C") groups started out at approximately the same point—due to randomization—with an average baseline test score of 85 (16th percentile).[36] This is similar to the actual baseline test scores seen for students in this study (see Tables II.2 through II.7).

Figure IV.1

Hypothetical Example of Gains and Impact



levels of reading ability. However, the students in this example (and in the actual study) began reading below grade level, indicated by standard scores less than 100. For such students, positive gains indicate the amount by which the students at least partially "caught up" to the average student in their grade. Negative gains indicate the amount by which the students fell further behind.

The impact shows the value added by the intervention; that is, the gain above that achieved by the control group. In other words, the impact is the amount that the interventions increased students' test scores relative to the control group. Because of random assignment, the intervention and control groups started out at the same place (85, in this example), and thus the impact can be calculated by comparing either the post-test scores for the intervention and control groups or the test score gains for the intervention and control groups. Using the post-test scores, the impact in Figure IV.1 is 95-89=6 (T_2-C_2). Alternatively, using gain scores, the impact in Figure IV.1 is (95-85)-(89-85)=10-4=6 ((T_2-T_1)-(C_2-C_1)). Thus, the intervention in this example raised students' test scores 6 points higher than they would have been without the intervention.

The change ("gain") in the intervention group students' average test scores between the beginning of the intervention year and the end of the following year can be calculated by adding the control group gain and the impact, as illustrated in Figure IV.1. If the control group students' average score increased between the beginning of the intervention year and end of the following school year and there is a positive impact, then the treatment group gain will also be positive, as in Figure IV.1, where the treatment group gain is 10 points. However, if the control group students' scores decreased during this period, then the intervention group may also experience a negative gain, even if the impact is positive. Depending on the relative magnitudes of the control group gain and the impact, a negative control group gain combined with a positive impact may imply that the intervention group students held their ground (or improved) while the control group declined, or may imply that the intervention group experienced a negative gain as well.

C. CONTEXT OF THE IMPACTS

We now consider our empirical findings pertaining to the three elements of the broader context for this evaluation: (1) where the students began in terms of reading ability just prior to the start of the interventions, (2) how much improvement the students would have had in the absence of the interventions, and (3) the amount of the interventions that treatment and control students actually received. Indicating where students began, the first column of Table IV.1 shows the baseline test scores of students on the tests we administered for this evaluation.[37] (Tables appear at the end of this chapter.[38]) The average baseline test scores are all below average (less than 100)—ranging from a low of 81 (10th percentile) for the Phonemic Decoding Efficiency test in the fifth-grade cohort to a high of 93 (32nd percentile) for Word Attack and Passage Comprehension in the fifth-grade cohort.

These estimates confirm that many students in our evaluation are not as severely impaired as many of the students studied in previous small-scale assessments of intensive reading interventions (see the review by Torgesen 2005). However, based upon teacher assessment and screening and baseline test

[37] As noted above, we present separate estimates for students in the third- and fifth-grade cohorts. At baseline, these students were in the third and fifth grades, respectively. When we administered the seven reading tests for estimating impacts in this report, most of the students were in the fourth and sixth grades.

[38] In addition to seven reading tests, we administered the spelling and calculation subtests from the Woodcock-Johnson III Tests of Achievement. Estimated impacts on scores for these tests are presented in Appendix E.

performance, the typical student in our evaluation is struggling with basic reading skills. That student, along with a substantial fraction of the broad range of students included in our sample are among those often targeted by providers and school districts for the types of interventions that we are evaluating. Such targeting is a response to both the needs of these students and the fact that except perhaps in the largest urban school districts, most schools would have only a small number of students in each grade who are as severely impaired as the students included in some previous studies. While it is important to assess the effectiveness of interventions for these more severely impaired students, the results obtained might not pertain to broader groups of struggling readers that include less severely impaired students. Hence, we have drawn our sample from regular elementary schools and included students with a relatively wide range of reading difficulties.

When we assess the improvement that students had achieved in the absence of the interventions, as measured by the tests that we administered for this evaluation, we see a mix of positive and negative gains among the control students in the third-grade cohort and mostly positive gains among students in the fifth-grade cohort, as presented in Table IV.1. In the third-grade pooled (ABCD) control group cohort, students typically showed little change, with gains between about –1 and 1 standard score points, but there was one larger positive gain and a larger negative gain. The negative gain on the Group Reading Assessment and Diagnostic Evaluation (GRADE) test suggests that the average student in the study lost ground relative to other students on this reading comprehension test between the start of the intervention year and the end of the following year. That is, if third-grade students selected for the study had not participated in an intervention, we would expect them, on average, to lose ground in their ability to extract meaning from text, as measured by the GRADE test. Among the fifth-grade cohort, the gains were generally positive, ranging from 1 to 5 standard score points for the interventions combined. The exceptions for fifth-grade control students were the reading comprehension tests, which showed almost no change (Passage Comprehension) or some negative gain (GRADE).

The positive gains experienced by the control students as measured by some of these tests, indicate that these students' reading ability improved in some dimensions between the beginning of the intervention year and the end of the following year, relative to the normal growth expected during this time. A

typical instruction, which included regular classroom instruction and often included other services, such as another pull out program.

As discussed in Chapter II, students in the intervention and control groups of the third-grade cohort received about the same amount of reading instruction during both the intervention year and the following year, although the intervention students received more small group and less large group instruction during the intervention year. For the fifth-grade cohort, the interventions not only shifted instruction from large to small groups during the intervention year but also increased the total amount of reading instruction received by students in the interventions. This latter increase was also observed during the next school year, although the available data do not reveal why. When we examine the impacts of the interventions on test scores from the seven reading tests that we administered for our second follow-up, we are comparing the effects of one year with the interventions followed by one year without for the intervention group students to the effects of two years without the interventions for the control group students.

The impacts presented in Tables IV.1 to IV.10, which are the ITT impacts, show the effects of students being given the opportunity to receive—from November into May during one academic year—a little less than one hour of intensive reading instruction per day, implemented as a pull out program from their usual classrooms, where they might have received

correction, and a more powerful adjustment developed by Benjamini and Hochberg (1995) that is particularly relevant for this study, where the interest is in assessing the impact of an intervention on multiple outcomes. The results in Appendix D show that adjustments for multiple comparisons do not affect the general conclusions of this report.

D. IMPACTS FOR STUDENTS IN THE THIRD-GRADE COHORT

Combined, the four interventions improved the phonemic decoding skills of the third-grade cohort one year after the intervention, raising Word Attack scores by approximately 5 standard score points (effect size 0.36)[39] and Phonemic Decoding Efficiency scores by approximately 4 points (effect size 0.26), as seen in Tables IV.1 and IV.11. These impacts for the pooled interventions (ABCD) suggest that being assigned to one of the reading interventions moved students in the interventions up the distribution of phonemic decoding ability approximately 5 to 10 percentile points more than they would have gained had they not been in one of the interventions.[40] The impacts of the four interventions combined are not the impacts of implementing the four interventions simultaneously. Rather, a combined impact can be interpreted as the impact of providing a struggling reader in third or fifth grade with the opportunity to receive a substantial amount of instruction in small groups with reasonably well-trained teachers, although as noted elsewhere, the content and instructional focus across the four interventions varied considerably. Such an impact is of greatest relevance to federal and state policymakers who can support broad programmatic approaches to instruction but cannot generally endorse specific products. In contrast, school district and school administrators must select specific products. For that purpose, the impact of being randomly assigned to an individual intervention—as modified or partially implemented for this study— is mo[(D. bs murt lntihis s57TEMC /P <</MCID 6 >0001 Tc -0.0001 T14)gs(dypfie)bl2211.5 276.6 34

The foulf 74.50 12.001 3Sfe1 3rven[72 s 1 3co1 3mbinTypa1 3Sd t1 3.50three 1 3word-level in1 3terv1 3en[72 s c1 3omb

F. IMPACTS FOR SUBGROUPS OF THE THIRD- AND FIFTH-GRADE COHORTS

Three of the four interventions—Spell Read, Wilson Reading, and Corrective Reading—focus on improving students' word-level reading skills. To examine whether the impacts of these interventions and the fourth intervention—Failure Free—were greater for students who began the interventions with more significant impairments in their word-level reading skills (specifically their phonemic decoding skills), we formed subgroups of students based on their entering scores on the Word Attack subtest. Students who began the study with lower scores on Word Attack were further subdivided into those who entered the study with lower or higher scores on the Peabody Picture Vocabulary Test. Since broad vocabulary is one of the significant factors that contribute to performance on measures of reading comprehension (Stahl 1998), it is of interest to determine whether the impact of the interventions varied among students with different entering scores on this dimension. In addition, because the No Child Left Behind legislation has increased funding for and attention to Title I schools, which by definition have high proportions of low-income students, we also examined the impacts of the interventions on students who qualified for free or reduced-price school lunch to determine if the interventions were particularly effective for that group.

The study was not designed to estimate the impacts of the individual interventions on subgroups of students, and thus did not enroll sufficient numbers of students to obtain precise estimates of such impacts. For this reason, we focus on the impacts of the four interventions combined and the three word-level interventions combined. The full subgroup results—including the estimated impacts of the individual interventions on subgroups of students—are presented in Tables IV.2 through IV.10, with effect sizes shown in Tables IV.12 through IV.20.

All of the tables of subgroup results contain two types of significance tests. One significance test is used to assess whether the impact for that subgroup is statistically different from zero, as indicated by an asterisk. That is, within a subgroup—for example, students in the third-grade cohort with baseline Word

1. Students with Relatively Low or High Word Attack Scores at Baseline

The first subgroup examined is students who entered the study with relatively low scores in phonemic decoding—specifically, Word Attack test scores below the 30th percentile. Approximately half of the students in each cohort had relatively low baseline scores on this test. Although the overall average score on the Word Attack test for this subgroup is still substantially higher than has been reported in many earlier intervention studies of substantially more impaired students at the same ages, there were no students in this group with average or above average scores in phonemic decoding before the interventions began.

Several of the impacts for students with low baseline Word Attack scores were similar to those for the full sample of students (see Table IV.2). Among students with low Word Attack scores in the third-grade cohort, the four interventions combined and the three word-level interventions combined had positive impacts on both measures of phonemic decoding, as was seen for the entire third-grade cohort. Likewise, the four interventions combined and the three word-level interventions combined improved scores on the measure of reading accuracy (Word Identification) and had no impact on the number of correct words read per minute on the oral reading passages (AIMSweb) or on reading comprehension according to the GRADE measure for all the students in the third-grade cohort and for those with low Word Attack scores. However, while the four interventions combined and the three word-level interventions combined also improved scores on the Sight Word Efficiency fluency test for the third-grade cohort as a whole, they did not improve scores on this test for students with low Word Attack scores. Finally, the interventions combined improved scores on the Passage Comprehension test for all students in the third-grade cohort, but not for students with low Word Attack scores.

Within the fifth-grade cohort, a few more impacts are seen for students with low Word Attack scores than for all students.[43] Along with raising the Word Attack test scores and lowering AIMSweb test scores, the four interventions combined improved scores on the Sight Word Efficiency test and the GRADE test for students with low Word Attack scores. The three word-level interventions improved scores for the entire fifth-grade cohort and for the fifth-grade cohort with low Word Attack scores on the phonemic decoding tests, and negatively impacted AIMSweb test scores, while also improving scores on the GRADE test for students with low Word Attack scores but not all students.

For at least several of the reading measures, impacts might seem smaller for students with low Word Attack scores in the third-grade cohort and larger for such students in the fifth-grade cohort in comparison with all students in those cohorts. However, differences in impacts are significant in only a few instances. Thus, we cannot conclude that low scores on the Word Attack test prior to the interventions made a reliable and consistent difference in the size of impacts obtained. Likewise, relatively high scores on the Word Attack test prior to the interventions do not reliably and consistently affect the size of impacts (see Table IV.3).

(continued)
estimates of such impacts. In fact, based on findings from previous studies, this evaluation was designed to detect fairly large impacts—even for all eligible students in a grade cohort—and not to estimate small impacts precisely.

[43] Some of the impacts, however, are no longer significant when we adjust for multiple comparisons.

2. Students with Relatively Low or High Vocabulary at Baseline

The impacts of the interventions may vary by students' broad vocabulary level. Therefore, we examined impacts for students with relatively high or relatively low verbal ability according to the Peabody Picture Vocabulary Test-Revised (selecting scores above or below the 30th percentile, respectively).[44] Forty-five percent of students in the third-grade cohort and 49 percent of students in the fifth-grade cohort had relatively low scores on this test. As described in Chapter II, we used this test in screening students for eligibility.

Although fewer impacts are seen with the interventions combined for the third-grade cohort with low Peabody Picture Vocabulary test scores than were seen for the entire third-grade cohort, only the difference in impacts for the AIMSweb test is statistically significant (see Table IV.4). In contrast, the four interventions combined had one more impact on third-grade students who began the year with relatively high Peabody Picture Vocabulary test scores (see Table IV.5), as compared to all of the third-grade cohort. Again, however, only the difference in the AIMSweb impact is statistically significant.

For the fifth-grade cohort, the patterns of impacts were generally similar for students with high and low Peabody Picture Vocabulary test scores. The four interventions combined increased GRADE scores for the students with low Peabody Picture Vocabulary test scores, but not for the students with high scores. However, the difference between the GRADE impacts was not significant.

3. Subgroups Defined Jointly by Baseline Phonemic Decoding and Vocabulary Scores

There was some expectation that the impacts of the interventions might be larger for students with low phonemic decoding ability but relatively high vocabulary, as this would create a sample that is more consistent with the way reading disabilities have been defined, and previous studies have found large impacts for students with severe disabilities (Lyon and Shaywitz 2003). Therefore, we examined impacts within subgroups defined by baseline Word Attack and screening Peabody Picture Vocabulary test scores. Each subgroup is approximately 25 percent of the full sample.

We did find seemingly different patterns of impacts across subgroups defined by these tests (see Tables IV.6 through IV.8), although many differences are not significant because sample sizes are small. The interventions combined improved more test scores for students in the third-grade cohort with relatively high scores on both the Word Attack test and the Peabody Picture Vocabulary test than for the other subgroups. In contrast, the interventions combined had fewer impacts for the fifth-grade cohort students with relatively high scores on both the Word Attack test and the Peabody Picture Vocabulary test than for the subgroups with lower Word Attack scores. The following is a summary of the impacts for the three groups of students of particular interest defined by these two tests:

- ***Students with Low Word Attack and Low Peabody Picture Vocabulary Test Scores.***[45] For the third-grade cohort students in this group, the four interventions combined and the three word-level interventions combined improved scores on the Word Attack test and

[44] The Peabody Picture Vocabulary Test, Third Edition (PPVT-III; Dunn and Dunn 1997) is a measure of receptive vocabulary in which the subject is required to select a picture that best depicts the verbal stimulus given by the examiner.

negatively impacted scores on the AIMSweb test. This is the only group within the third-grade cohort where the combined interventions had a negative impact on AIMSweb scores, an impact that is significantly different than the results for the entire third-grade cohort. For students in the fifth-grade cohort in this group, the four interventions combined had positive impacts on scores on both of the phonemic decoding tests, the Sight Word Efficiency test, and the Passage Comprehension test, but a negative impact on scores on the AIMSweb test.

- ***Students with Low Word Attack and High Peabody Picture Vocabulary Test Scores.***[46] For students in the third-grade cohort, the four interventions combined had positive impacts on scores on the Word Identification, AIMSweb, and Passage Comprehension tests. The three word-level interventions improved only the AIMSweb test scores. For students in the fifth-grade cohort, the four interventions combined and the three word-level interventions combined had impacts on both measures of phonemic decoding, the Sight Word Efficiency test, and the GRADE test.

- ***Students with High Word Attack and High Peabody Picture Vocabulary Test Scores***.[47] For students in the third-grade cohort, the four interventions combined and the three word-level interventions combined improved scores on both measures of phonemic decoding and two1 [72 744a5ougWores on c

2. Impacts for the Fifth-Grade cohort

Similar to the results for the third-grade cohort, most impacts on reading scores for the fifth-grade cohort are negative. The four interventions combined lowered scores by 27 scaled points (effect size -0.11), although Failure Free Reading was, again the only individual intervention to show a significant impact. Among fifth graders with relatively high baseline Word Attack or relatively high screening vocabulary scores or both relatively high baseline Word Attack and screening vocabulary scores, the four interventions combined lowered reading scores. In addition, the four interventions combined lowered scores for students eligible for free or reduced-price lunch by 74 scaled points (effect size -0.31).

The four interventions combined and the three word-level interventions combined lowered math scores for students in the fifth-grade cohort by 29 and 34 scaled points (effect sizes -0.12 and -0.14), respectively, primarily due to the negative impacts of Wilson Reading and Corrective Reading. The four interventions combined also lowered scores for students with relatively high baseline Word Attack scores and for students with low screening vocabulary scores.

I. SUMMARY OF KEY FINDINGS

Although many estimates have been provided in this report, our key findings are as follows:[54]

1. ***The interventions improved some reading skills.*** For students in the third-grade cohort ,the tour interv
resding saocuadcyand Cfunntc, din5(vngcombr)6(

Figure IV.4

Gap Reduction for Third-Grade Cohort: Word Identification

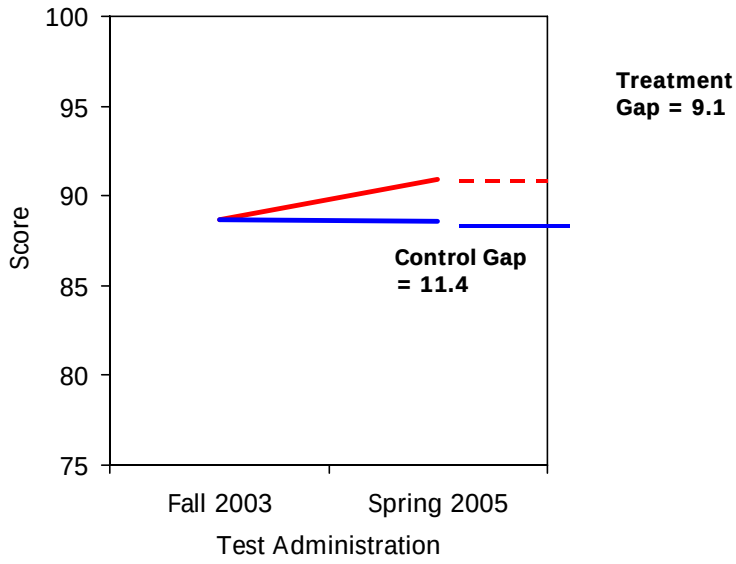


Figure IV.8

Gap Reduction for Fifth-Grade Cohort: Word Attack

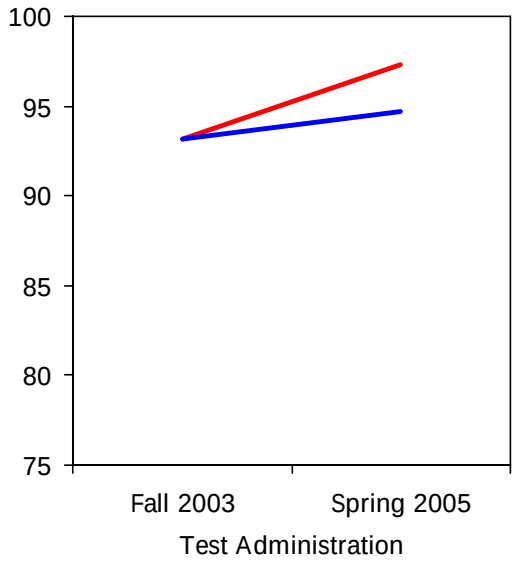


Figure IV.12

Gap Reduction for Fifth-Grade Cohort: Passage Comprehension

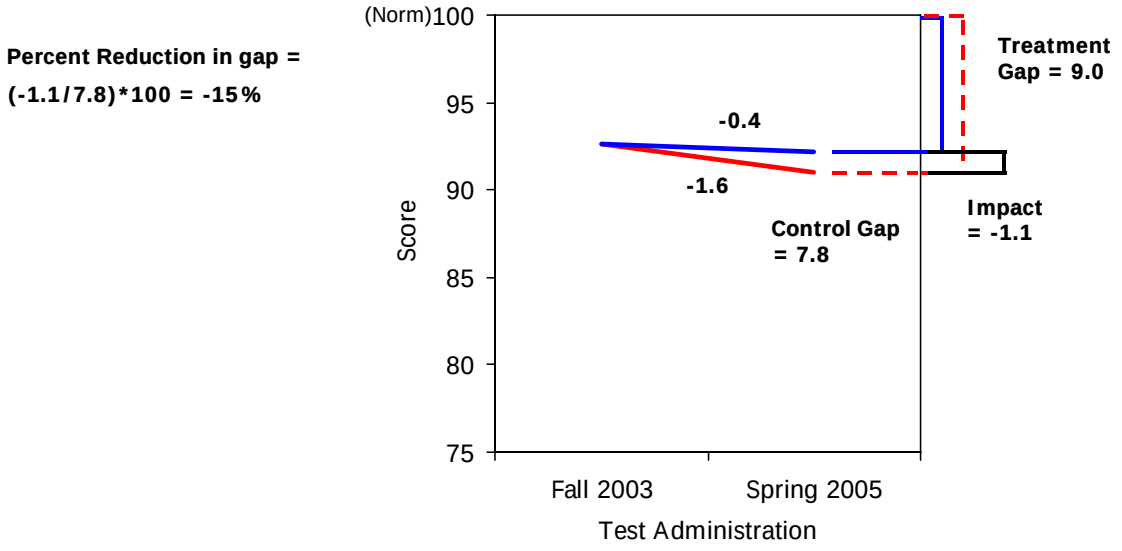


Figure IV.13

Gap Reduction for Fifth-Grade Cohort: GRADE

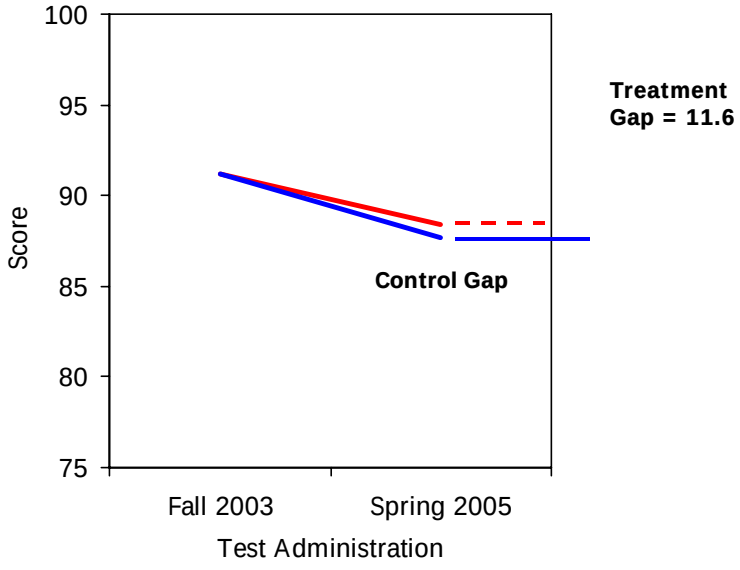


Table IV.3

Impacts for 3rd and 5th Grade Cohorts with High Baseline Word Attack Scores
One Year After the Intervention Year

		All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
Grade 3 Cohort	Baseline	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	100.9	-3.5	6.5 *	-2.2	5.6 *	-7.4	9.3 *	-2.5	3.7	-1.3	4.8 *	-2.9	8.3 *
TOWRE PDE	89.2	2.2	3.6 *	1.8	4.8 *	3.3	0.1	3.2	5.4 *	0.6	6.1 * #	1.5	2.8 *
Word Identification	92.2	0.4	3.0 *	0.6	3.0 *	-0.1	3.0 *	2.0	0.9	-2.1	4.4 *	1.8	3.8
TOWRE SWE	90.5	1.6	2.8 * #	1.7	2.7 * #	1.3	3.1 *	3.0	0.9	0.9	3.1 *	1.2	4.2 * #
AIMSweb	50.2	33.1	4.9 *	31.2	6.3 *	38.8	0.7	32.0	3.9	27.4	10.8 *	34.1	4.2
Passage Comprehension	96.8	-2.5	2.0 *	-1.7	0.6	-5.0	6.0 *	-2.1	1.3	-1.7	-0.8	-1.2	1.4
GRADE	89.4	-4.9	-0.5	-4.6	-2.4	-5.8	5.2 *	-10.9	4.9 #	-3.6	-10.5 * #	0.7	-1.4
Sample Size	159		159		126		33		44		40		42

		All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
Grade 5 Cohort	Baseline	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	101.6	0.1	0.8 #	1.3	1.2 #	-3.5	-0.2	2.1	-1.8 #	1.1	6.0 *	0.8	-0.8
TOWRE PDE	86.4	5.4	1.2	6.0	1.1	3.7	1.5	5.5	2.0	5.3	1.0	7.3	0.2
Word Identification	93.2	4.7	-1.8 *	5.1	-1.7	3.7	-2.2	4.4	-2.2	4.5	0.6	6.3	-3.3
TOWRE SWE	87.1	3.1	1.1	3.1	1.2	3.1	0.8	1.5	3.3 *	2.8	0.6	5.0	-0.4
AIMSweb	86.8	30.0	-1.5	29.6	-1.2	31.0	-2.2	25.7	3.1	26.3	0.5	37.0	-7.3
Passage Comprehension	96.0	-0.8	-2.1 *	-0.8	-1.9	-0.9	-2.8	-3.7	-2.5	-1.0	0.1	2.3	-3.2 *
GRADE	94.3	-3.4	-1.7 #	-4.7	-1.1 #	0.6	-3.4 *	-7.0	0.2	-6.2	-1.0 #	-0.9	-2.3 * #
Sample Size	205		205		135		70		41		46		48

Note: Raw scores were analyzed for the AIMSweb, and standard scores were analyzed for all other tests.

* 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 IV.4

Impacts for 3rd and 5th Grade Cohorts with Low Screening Peabody Picture Vocabulary Test Scores
One Year After the Intervention Year

		All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
Grade 3 Cohort	Baseline	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	92.0	0.0	4.0 *	1.9	3.0 *	-5.8	6.9 *	6.0	-0.7 #	-0.4	6.2 *	0.2	3.6 *
TOWRE PDE	85.3	1.6	3.6 *	2.4	3.9 *	-0.6	2.8	6.9	2.1	2.8	3.6 *	-2.5	6.1 * #
Word Identification	87.9	0.1	1.7	0.4	1.7	-0.9	1.7	2.8	-0.7	-1.7	3.9 *	0.1	2.0
TOWRE SWE	86.0	3.6	1.0	3.8	1.0	3.0	0.8	3.5	0.8	3.3	2.1	4.6	0.2
AIMSweb	39.0	38.0	-2.1 #	40.1	-6.3 #	31.7	10.4	36.0	-1.9	47.4	-16.2 #	36.8	-0.9
Passage Comprehension	89.9	-0.1	0.6	0.6	0.1	-2.4	2.1	0.3	0.5	1.4	-1.2	0.2	1.2
GRADE	83.4	-8.7	1.1	-9.2	1.9	-7.1	-1.5 #	-14.6	5.9	-11.2	3.3	-1.9	-3.3
Sample Size	147		147		110		37		37		42		31

		All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
Grade 5 Cohort	Baseline	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	91.4	1.9	2.4 *	2.0	3.9 *	1.7	-1.9	2.4	1.2	4.4	7.2 *	-0.8	3.2
TOWRE PDE	79.9	5.3	2.2 *	5.5	2.6 *	4.7	1.2	6.6	1.0 #	5.7	2.1	4.2	4.7
Word Identification	86.7	1.6	-0.2	2.3	-0.4	-0.7	0.5	1.7	-1.2 #	2.8	1.3	2.4	-1.5
TOWRE SWE	83.5	1.9	2.4 *	1.4	3.0 *	3.3	0.7	1.4	2.7 *	1.0	2.1	1.8	4.0 * #
AIMSweb	72.3	34.7	-6.6	32.6	-5.0	40.9	-11.4	28.1	-6.7	28.0	0.0	41.9	-8.3
Passage Comprehension	89.4	-2.5	1.6 #	-2.8	2.2 * #	-1.5	-0.4	-3.0	1.3	-3.0	3.3 * #	-2.5	2.0
GRADE	87.0	-3.9	3.5 *	-4.1	3.3 *	-3.4	4.0 * #	-3.5	-0.3	-8.1	5.1 *	-0.7	5.2 *
Sample Size	195		195		140		55		53		52		35

Note: Raw scores were analyzed for the AIMSweb, and standard scores were analyzed for all other tests.

* 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 IV.6

Impacts for 3rd and 5th Grade Cohorts with Low Baseline Word Attack and Low Screening Peabody Picture Vocabulary Test Scores
One Year After the Intervention Year

		All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
Grade 3 Cohort	Baseline	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	85.3	1.6	3.7 *	2.1	4.5 *	-0.1	1.0	3.6	4.2	3.2	3.5	-0.5	5.9 *
TOWRE PDE	82.0	3.0	1.3	4.1	1.1	-0.2	1.8	11.5	-3.7 #	-0.3	3.7	1.1	3.3
Word Identification	85.1	0.3	0.7	0.9	0.4	-1.6	1.4	3.6	-1.4	-0.3	1.3	-0.5	1.3
TOWRE SWE	82.3	5.5	-0.6 #	5.8	-0.6	4.4	-0.5	4.4	2.8	8.3	-4.2	4.8	-0.3
AIMSweb	31.5	40.9	-6.9 * #	42.8	-10.7 * #	35.0	4.8	36.8	-3.1	60.0	-35.3 * #	31.7	6.1
Passage Comprehension	85.4	2.1	0.6	3.0	0.0	-0.6	2.6	4.4	-1.9	2.0	1.1	2.6	0.7
GRADE	81.4	-8.1	-0.7	-9.4	1.4	-4.3	-6.8 #	-13.8	4.1	-9.7	4.2	-4.8	-4.2
Sample Size	81		81		55		26		24		14		17

		All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
Grade 5 Cohort	Baseline	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	83.4	4.7	2.7 *	4.3	4.2 *	5.9	-1.9	0.6	4.3 *	10.1	3.4 #	2.1	4.9 * #
TOWRE PDE	74.6	5.3	2.7 *	5.5	3.2 *	4.5	1.1	6.9	0.9	5.7	2.9	3.9	5.8
Word Identification	82.6	0.9	0.7	1.5	0.4	-1.0	1.4	4.4	-1.4	-0.2	1.3	0.1	1.3
TOWRE SWE	80.0	2.2	3.0 *	1.7	3.3	3.9	2.1	1.8	3.4 *	1.7	2.2	1.6	4.4
AIMSweb	63.1	34.3	-10.8 *	34.1	-12.2 *	34.7	-6.6	25.5	-11.2	31.5	-7.6	45.4	-17.7
Passage Comprehension	86.1	-2.6	3.9 * #	-3.3	5.1 * #	-0.5	0.1	-5.3	5.7 * #	-1.4	2.9	-3.3	6.7 * #
GRADE	84.3	-4.4	2.7	-4.1	1.8	-5.2	5.7 #	-3.6	-2.6	-8.9	4.5	0.1	3.3
Sample Size	107		107		77		30		34		25		18

Note: Raw scores were analyzed for the AIMSweb, and standard scores were analyzed for all other tests.

* 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.

Grade 3 Cohort	Baseline	Control Gain			Control Gain			Control Gain			Control Gain			Control Gain			Control Gain		
Word Attack	83.9	7.1	0.1	#	8.0	0.1		4.1	-0.2		5.7	3.0		4.3	3.5		14.2	-6.1	* #
TOWRE PDE	82.1	2.7	1.8		3.3	1.9	#	0.9	1.5		2.9	5.2	*	2.4	3.2		4.7	-2.7	#
Word Identification	85.5	0.5	2.1	*	0.4	2.3		0.7	1.4		1.4	2.2		-3.5	5.7	*	3.2	-0.9	
TOWRE SWE	83.3	4.3	1.3		4.3	1.2		4.4	1.5		4.0	0.1		0.9	5.6		7.9	-1.9	
AIMSweb	32.9	31.1	9.5	*	29.9	10.4	*	34.8	6.8	*	27.1	8.6		17.5	18.6		45.1	4.1	
Passage Comprehension	87.9	1.4	3.6	*	1.4	3.5		1.3	3.9	*	5.0	-2.0		-5.1	9.4		4.3	3.2	
GRADE	84.5	-7.4	2.7		-5.8	0.9		-12.3	7.9	* #	-5.3	-3.7		-14.7	4.8		2.7	1.6	
Sample Size	89		89			59			30			23			16			20	

Grade 5 Cohort	Baseline	Control Gain			Control Gain			Control Gain			Control Gain			Control Gain			Control Gain		
Word Attack	86.0	2.5	4.6	*	2.3	6.5	* #	3.2	-1.2		-0.1	7.2	*	1.9	10.7	*	5.1	1.7	
TOWRE PDE	76.5	3.6	3.5	*	2.5	5.5	* #	6.8	-2.5		3.1	7.3	* #	0.3	6.0	*	4.2	3.1	*
Word Identification	85.4	2.1	-0.3		11(-)12(3.)-()12(3.)-iTJ11.162 0Td[2.)-6(7)-4292(-)13(1.)-5(9)-360(0 Td[C)a24t1(b)]TJ12.516 0 T-5(5)]TJ9 0 Td[C[5)7(9M2(-)13(1.)-5(9 Td[)56													

Grade 3 Cohort	Baseline	Control Gain		Control Gain		Control Gain		Control Gain		Control Gain		Control Gain	
Word Attack	101.3	-3.1	5.7 *	-3.3	6.6 *	-2.8	3.1 *	-9.5	10.3 * #	-1.1	4.4 *	0.8	5.2 *
TOWRE PDE	89.1	2.5	2.5 *	2.1	3.7 *	3.7	-1.2	2.6	5.1 *	2.7	4.0 *	1.0	2.1
Word Identification	92.8	1.0	1.6 *	0.5	2.4 *	2.3	-0.7	2.0	1.0	-1.1	1.0	0.7	5.1 *
TOWRE SWE	90.5	2.4	2.5 *	2.5	2.6 *	2.0	2.3 *	4.9	0.1	3.0	2.6 *	-0.2	5.1 * #
AIMSweb	51.7	32.8	5.2	30.7	6.9	39.3	0.0	29.4	4.3	24.3	15.9 *	38.3	0.4
Passage Comprehension	97.7	-3.0	2.5 *	-2.0	0.8	-5.9	7.5 *	-1.8	0.2	-3.5	0.6	-0.9	1.6
GRADE	91.9	-6.8	1.0	-7.5	1.0	-4.8	1.1	-13.3	6.5	-7.5	-4.7	-1.7	1.3
Sample Size	93		93		71		22		31		12		28

Grade 5 Cohort	Baseline	Control Gain		Control Gain		Control Gain		Control Gain		Control Gain		Control Gain	
Word Attack	101.8	0.2	1.2	1.6	1.7	-3.9	-0.1	-0.9	2.1	3.0	4.3	2.7	-1.4
TOWRE PDE	86.4	5.0	2.3 *	5.5	2.7 *	3.6	1.1	5.0	3.7	3.8	3.1	7.6	1.3
Word Identification	94.3	4.6	-1.2	5.0	-1.4	3.5	-0.5	4.1	-2.6 *	6.4	-0.2	4.4	-1.5
TOWRE SWE	86.7	4.1	1.5	4.5	1.4	3.0	1.9	4.2	1.2	3.6	4.5	5.7	-1.6
AIMSweb	89.1	29.1	-2.6	29.5	-3.2	28.2	-0.7	23.6	1.0	24.0	4.0	40.8	-14.5 * #
Passage Comprehension	97.7	-1.1	-0.9	-1.1	-0.3	-1.1	-2.8	-5.1	-0.3	0.0	0.5	1.9	-1.0
GRADE	97.1	-3.8	-4.3 * #	-4.8	-4.4 * #	-1.0	-4.0 *	-7.5	-5.4	-4.7	-2.6	-2.1	-5.4 * #
Sample Size	117		117		72		45		22		19		31

Note: Raw scores were analyzed for the AIMSweb, and

Table IV.9

Impacts for 3rd and 5th Grade Cohorts Eligible for Free or Reduced-Price School Lunch
One Year After the Intervention Year

		All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
Grade 3 Cohort	Baseline	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	91.9	2.0	3.7 *	3.3	2.7 *	-2.0	6.7 *	4.4	1.5	2.5	4.2 *	2.9	2.3 *
TOWRE PDE	85.1	3.3	2.8 * #	3.9	3.1 * #	1.3	2.0	4.5	5.4 *	3.8	3.2 * #	3.5	0.7 #
Word Identification	87.9	0.6	3.2	0.9	3.4	-0.1	2.5	4.1	-0.3	-1.9	5.6 *	0.5	5.1 *
TOWRE SWE	85.7	3.1	2.1 *	3.6	1.5	1.7	3.6 *	2.7	1.6	3.3	2.8 *	4.9	0.1 *
AIMSweb	38.7	36.6	2.8	36.0	1.7	38.5	5.9	32.6	3.6	34.1	0.3	41.4	1.2
Passage Comprehension	90.2	0.9	-0.1 #	1.0	-0.1	0.6	-0.2 #	1.7	-0.4	0.1	0.2	1.1	0.0
GRADE	84.4	-7.0	1.0	-5.8	-0.5	-10.6	5.6	-8.4	0.1	-10.3	1.5	1.5	-3.1
Sample Size	190		190		144		46		53		46		45

		All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
Grade 5 Cohort	Baseline	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	92.5	2.2	2.9 *	3.0	3.8 *	0.0	0.0	1.5	3.5	5.2	6.1 *	2.3	1.8
TOWRE PDE	80.1	5.8	1.2	6.1	1.6	5.0	0.0	6.9	0.6	6.9	0.2	4.5	3.9 *
Word Identification	87.7	2.0	-0.1	2.4	0.1	0.6	-0.9	1.0	0.6	4.0	0.4	2.3	-0.7
TOWRE SWE	83.0	2.0	2.6 *	2.0	3.0 *	2.2	1.3	1.4	4.0 *	1.9	1.5	2.7	3.6 * #
AIMSweb	72.4	25.9	1.7	24.6	4.6 #	29.7	-7.0	20.9	3.7 #	23.7	3.9	29.2	6.1 #
Passage Comprehension	90.4	-0.5	-1.2	-1.1	-0.5	1.1	-3.2	-1.1	-2.1	-1.9	1.1	-0.3	-0.5
GRADE	88.0	-0.9	-3.3 *	-2.3	-2.9 *	3.3	-4.5 *	1.6	-9.4 * #	-5.6	0.4	-2.8	0.3
Sample Size	220		220		150		70		50		56		44

Note: Raw scores were analyzed for the AIMSweb, and standard scores were analyzed for all other tests.

* 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.

Grade 3 Cohort	Baseline	Control Gain		Control Gain		Control Gain		Control Gain		Control Gain		Control Gain	
Word Attack	93.2	-2.1	6.1 *	-2.3	7.7 *	-1.8	1.0	-3.0	6.6 *	-3.6	9.7	-0.2	7.0 *
TOWRE PDE	86.0	-2.8	7.8 * #	-4.0	10.3 * #	1.0	0.4	2.2	5.3 *	-10.2	15.8 * #	-4.1	9.9 * #
Word Identification	89.8	-0.2	2.4 *	-0.1	2.2 *	-0.6	2.9 *	1.1	0.9	-2.5	3.9	1.2	1.8
TOWRE SWE	87.8	3.3	2.1 *	3.3	2.3	3.1	1.6 *	4.9	-0.1	1.4	4.1	3.7	3.0
AIMSweb	44.1	27.1	10.9 *	25.4	12.3 *	32.4	6.6 *	30.7	1.9	6.3	29.3 *	39.1	5.7
Passage Comprehension GRADE	93.5	-5.0	6.0 * #	-5.0	5.5	-4.9	7.7 * #	0.3	-0.5	-16.5	15.7	1.2	1.3

Grade 3 Cohort						
Word Attack	0.36 *	0.37 *	0.33 *	0.36 *	0.38 *	0.35 *
TOWRE PDE	0.26 *	0.32 *	0.08	0.33 *	0.28 *	0.37 *
Word Identification	0.15 *	0.17 *	0.12	0.05	0.28 *	0.17 *
TOWRE SWE	0.11 *	0.11 *	0.13 *	0.06	0.17 *	0.09
AIMSweb	0.14	0.11	0.20 *	0.15 *	0.09	0.10
Passage Comprehension	0.14 *	0.09	0.29 *	0.01	0.23	0.02
GRADE	0.06	0.02	0.19	0.14	0.00	-0.07

Grade 5 Cohort						
Word Attack	0.18 * #	0.26 *	-0.05 #	0.23	0.52 *	0.01 #
TOWRE PDE	0.11	0.16 *	-0.02	0.21	0.17	0.09
Word Identification	-0.04 #	-0.04 #	-0.04 #	0.01	0.00 #	-0.13 #
TOWRE SWE	0.09	0.09	0.10	0.23 *	0.08	-0.03
AIMSweb	-0.08 * #	-0.08 * #	-0.09 #	-0.07 #	-0.06	-0.11
Passage Comprehension	-0.08 #	-0.05	-0.17 #	-0.06	0.06	-0.14
GRADE	0.05	0.08	-0.06	-0.07	0.31	0.00

Note: Population standard deviation = 15 for all tests except AIMSweb.

Grade 3 Cohort

Word Attack

Grade 3 Cohort

Word Attack

Grade 3 Cohort						
Word Attack	0.27 *	0.20 *	0.46 *	-0.05 #	0.41 *	0.24 *
TOWRE PDE	0.24 *	0.26 *	0.19	0.14	0.24 *	0.40 * #
Word Identification	0.11	0.11	0.11	-0.05	0.26 *	0.13
TOWRE SWE	0.07	0.07	0.06	0.05	0.14	0.01

Grade 3 Cohort						
Word Attack	0.29 *	0.34 *	0.13	0.51 * #	0.38	0.14 *
TOWRE PDE	0.18 *	0.20 *	0.13	0.44 *	0.18	-0.01 #

Table IV.17

Effect Sizes for 3rd and 5th Grade Cohorts With Low Baseline Word Attack and High Peabody Picture Vocabulary Test Scores
One Year After the Intervention Year

Grade 3 Cohort	All Interventions ABCD Effect Size	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read B Impact	Wilson Reading C Impact	Corrective Reading D Impact
Word Attack	0.00 #	0.01	-0.02	0.20	0.23	-0.40 * #
TOWRE PDE	0.12	0.13 #	0.10	0.35 *	0.21	-0.18 #
Word Identification	0.14 *	0.15	0.10	0.14	0.38 *	-0.06
TOWRE SWE	0.09	0.08	0.10	0.01	0.37	-0.13
AIMSweb	0.24 *	0.27 *	0.17 *	0.22	0.47	0.11
Passage Comprehension	0.24 *	0.24	0.26 *	-0.13	0.63	0.21
GRADE	0.18	0.06	0.53 * #	-0.24	0.32	0.11

Grade 5 Cohort	All Interventions ABCD Effect Size	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read B Impact	Wilson Reading C Impact	Corrective Reading D Impact
Word Attack	0.31 *	0.43 * #	-0.08	0.48 *	0.71 *	0.11
TOWRE PDE	0.23 *	0.36 * #	-0.16	0.48 * #	0.40 *	0.21 *
Word Identification	-0.02	0.02	-0.12 * #	0.15 #	0.00	-0.09
TOWRE SWE	0.17 *	0.19 *	0.14	0.29 *	0.22	0.05
AIMSweb	-0.16	-0.13	-0.26	-0.04	-0.31 * #	-0.03
Passage Comprehension	-0.04	-0.04	-0.06	-0.26	0.22	-0.07
GRADE	0.43 * #	0.61 * #	-0.09	0.34	0.99 * #	0.49

Note: Population standard deviation = 15 for all tests except AIMSweb.
AIMSweb SD (Fall) 3rd grade = 39.2; AIMSweb SD (Fall) 5th grade = 47

Note: Raw scores were analyzed for the AIMSweb, and standard scores were analyzed for all other tests.

* 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.

Grade 3 Cohort						
Word Attack	0.38 *	0.44 *	0.20 *	0.69 * #	0.29 *	0.35 *

Table IV.19

Effect Sizes for 3rd and 5th Grade Cohorts Eligible for Free or Reduced-Price School Lunch
One Year After the Intervention Year

Grade 3 Cohort	All Interventions ABCD Effect Size	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read B Impact	Wilson Reading C Impact	Corrective Reading D Impact
Word Attack	0.24 *	0.18 *	0.45 *	0.10	0.28 *	0.16 *
TOWRE PDE	0.19 * #	0.21 * #	0.13	0.36 *	0.21 * #	0.05 #
Word Identification	0.21	0.23	0.17	-0.02	0.37 *	0.34 *
TOWRE SWE	0.14 *	0.10	0.24 *	0.11	0.19 *	0.01 *
AIMSweb	0.07	0.04	0.15	0.09	0.01	0.03
Passage Comprehension	-0.01 #	0.00	-0.02 #	-0.03	0.01	0.00
GRADE	0.07	-0.03	0.37	0.01	0.10	-0.21

Grade 5 Cohort	All Interventions ABCD Effect Size	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read B Impact	Wilson Reading C Impact	Corrective Reading D Impact
Word Attack	0.19 *	0.25 *	0.00	0.23	0.41 *	0.12
TOWRE PDE	0.08	0.10	0.00	0.04	0.01	0.26 *
Word Identification	-0.01	0.01	-0.06	0.04	0.03	-0.05
TOWRE SWE	0.17 *	0.20 *	0.09	0.27 *	0.10	0.24 * #
AIMSweb	0.04	0.10 #	-0.15	0.08 #	0.08	0.13 #
Passage Comprehension	-0.08	-0.03	-0.21	-0.14	0.07	-0.03
GRADE	-0.22 *	-0.19 *	-0.30 *	-0.63 * #	0.03	0.02

Note: Population standard deviation = 15 for all tests except AIMSweb.
AIMSweb SD (Fall) 3rd grade = 39.2; AIMSweb SD (Fall) 5th grade = 47

Note: Raw scores were analyzed for the AIMSweb, and standard scores were analyzed for all other tests.

* 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 IV.20

Effect Sizes for 3rd and 5th Grade Cohorts Not Eligible for Free or Reduced-Price School Lunch
One Year After the Intervention Year

Grade 3 Cohort	All Interventions ABCD Effect Size	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read B Impact	Wilson Reading C Impact	Corrective Reading D Impact
Word Attack	0.40 *	0.52 *	0.07	0.44 *	0.65	0.47 *
TOWRE PDE	0.52 * #	0.69 * #	0.03	0.35 *	1.05 * #	0.66 * #
Word Identification	0.16 *	0.15 *	0.19 *	0.06	0.26	0.12
TOWRE SWE	0.14 *	0.16	0.10 *	0.00	0.27	0.20
AIMSweb	0.28 *	0.31 *	0.17 *	0.05	0.75 *	0.14
Passage Comprehension	0.40 * #	0.37	0.51 * #	-0.03	1.04	0.09
GRADE	0.02	0.01	0.05	0.17	-0.27	0.12

Grade 5 Cohort	All Interventions ABCD Effect Size	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read B Impact	Wilson Reading C Impact	Corrective Reading D Impact
Word Attack	0.20 *	0.30 *	-0.07	0.31 *	0.67 *	-0.09
TOWRE PDE	0.14 *	0.18 *	0.01	0.30 *	0.27 *	-0.02
Word Identification	-0.06	-0.08	0.00	0.01	-0.02	-0.23
TOWRE SWE	0.10	0.09	0.13	0.23	0.16 *	-0.11 #
AIMSweb	-0.09	-0.12 * #	0.00	-0.14 * #	-0.06	-0.16 * #
Passage Comprehension	-0.10	-0.09	-0.12	-0.05	-0.11	-0.12
GRADE	0.06	0.09	-0.01	0.17 #	0.20	-0.12

Note: Population standard deviation = 15 for all tests except AIMSweb.
AIMSweb SD (Fall) 3rd grade = 39.2; AIMSweb SD (Fall) 5th grade = 47

Note: Raw scores were analyzed for the AIMSweb, and standard scores were analyzed for all other tests.

* 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 IV.22

Impacts for 3rd and 5th Grade Cohorts
Late March/Early April of the Intervention Year

Grade 3 Cohort	All Interventions ABCD Impact	Word-level Interventions BCD Impact	Failure Free Reading CA Impact	Spell Read DB Impact	Wilson Reading	Corrective Reading
PSSA Reading	-15.6	-3.8	-51.1 *	-39.9	52.5	-23.8
PSSA Math	20.2	14.2	38.4	-15.5	56.6 *	1.4
Sample Size	329	240	89	92	71	77

Grade 5 Cohort	All Interventions ABCD Impact	Word-level Interventions BCDB Impact	Failure Free Reading C Impact	Spell Read D Impact	Wilson Reading	Corrective Reading
PSSA Reading	-27.3 *	-25.3	-33.4 *	-30.0	-23.8	-22.1
PSSA Math	-28.8 * #	-34.0 * #	-13.4	-20.1	-56.4 * #	-25.4 *
Sample Size	408	280	128	102	92	86

* Impact is statistically significant at the 0.05 level.
# Impact is statistically different from the 3rd grade cohort impact at the 0.05 level.

Table IV.23

Grade 3 Cohort	All Interventions ABCD Impact	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read	Wilson Reading C Impact	Corrective Reading D Impact
PSSA Reading	-1.9	25.9 #	-85.3 *	-13.6	59.7	31.4
PSSA Math	46.1 * #	51.3 * #	30.7	-32.9	120.7 *	65.9 * #
Sample Size	168	113	55	47	31	35

Grade 5 Cohort					Wilson Reading	Corrective Reading
PSSA Reading	-29.3	-23.8	-45.8	-39.5	-43.7	11.7
PSSA Math	-10.8	-19.3	14.8	-16.7	-28.0	-13.2
Sample Size	200	142	58	60	44	38

* 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 IV.24

Impacts for 3rd and 5th Grade Cohorts with High Baseline Word Attack Scores
Late March/Early April of the Intervention Year

Grade 3 Cohort	All Interventions ABCD Impact	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read B Impact	Wilson Reading C Impact	Corrective Reading D Impact
PSSA Reading	-21.2	-23.1 #	-15.6	-63.7 *	25.0	-30.5
PSSA Math	-3.1 #	-13.7 #	28.4	-7.5	32.3	-65.8 * #
Sample Size	208	138	70	42	48	48

Grade 5 Cohort	All Interventions ABCD Impact	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read B Impact	Wilson Reading C Impact	Corrective Reading D Impact
PSSA Reading	-36.3 *	-40.6	-23.3	-37.9	-17.5	-66.3
PSSA Math	-39.0 *	-42.8 *	-27.7	-12.2	-53.8	-62.2
Sample Size	203	148	55	57	55	36

* 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 IV.25

Impacts for 3rd and 5th Grade Cohorts with Low Screening Peabody Picture Vocabulary Test Scores
Late March/Early April of the Intervention Year

Grade 3 Cohort	All Interventions ABCD Impact	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read B Impact	Wilson Reading C Impact	Corrective Reading D Impact
PSSA Reading	25.1	37.3	-11.4	21.0	59.3	31.7
PSSA Math	83.9	81.3	91.7	76.4 #	126.9 *	40.8
Sample Size	143	108	35	36	43	29

Grade 5 Cohort	All Interventions ABCD Impact	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read B Impact	Wilson Reading C Impact	Corrective Reading D Impact
PSSA Reading	2.0	2.2	1.6	-30.2	-56.4 *	92.9 * #
PSSA Math	-33.8 *	-40.2	-14.6	-42.9	-43.8	-34.0
Sample Size	203	148	55	57	55	36

* 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 IV.26

Impacts for 3rd and 5th Grade Cohorts with High Screening Peabody Picture Vocabulary Test Scores
Late March/Early April of the Intervention Year

Grade 3 Cohort	All Interventions ABCD Impact	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read B Impact	Wilson Reading C Impact	Corrective Reading D Impact
PSSA Reading	0.9	5.6	-13.0	-25.4	38.6	3.6
PSSA Math	6.7	13.5	-13.9	-41.3 #	75.1	6.8
Sample Size	186	132	54	56	28	48

Grade 5 Cohort	All Interventions ABCD Impact	Word-level Interventions BCD Impact	Failure Free Reading A Impact	Spell Read B Impact	Wilson Reading C Impact	Corrective Reading D Impact
PSSA Reading	-32.6 *	-26.7	-50.5 *	-3.6	-9.6	-66.8 * #
PSSA Math	-17.8	-22.4	-4.2	26.7	-82.0 *	-11.8
Sample Size	205	132	73	45	37	50

* 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.

Grade 3 Cohort						
PSSA Reading	57.5	82.9 #	-18.8	34.8	101.3	112.6 * #
PSSA Math	131.9 * #	139.0 * #	110.8	125.0	172.1 *	119.8
Sample Size	77	53	24	23	15	15
Grade 5 Cohort						
PSSA Reading	-0.6	8.3	-26.8	-28.6	-94.3 *	147.4 * #
PSSA Math	-22.7	-26.2	-11.9	-32.6	-61.6	15.5
Sa						

Table IV.30

					C	
Grade 3 Cohort						
PSSA Reading	-66.3 * #	-45.2 #	-129.5 * #	-57.7	-25.5	-52.5 #
PSSA Math	10.1	-0.8	42.9	-31.7	45.1 * #	-15.7
Sample Size	190	144	46	52	47	45

					Wilson Reading	Corrective Reading
Grade 5 Cohort						
PSSA Reading	-74.1 * #	-67.3 *	-94.5 * #	-110.7 * #	-69.8 *	-21.5
PSSA Math	-35.3	-36.6	-31.3	-58.7	-30.3	-20.8
Sample Size	230	160	70	53	61	46

* 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 IV.31

Impacts for 3rd and 5th Grade Cohorts Not Eligible for Free or Reduced Price School Lunch
Late March/Early April of the Intervention Year

Grade 3 Cohort	All Interventions ABCD Impact		Word-level Interventions BCD Impact		Failure Free Reading A Impact		Spell Read B Impact	Wilson Reading C Impact		Corrective Reading D Impact	
PSSA Reading	43.4	#	47.6	#	30.7	#	9.6	90.5		42.7	#
PSSA Math	-18.3		-34.3		29.8		-12.9	-100.3	#	10.2	
Sample Size	139		96		43		40	24		32	

Grade 5 Cohort	All Interventions ABCD Impact		Word-level Interventions BCD Impact		Failure Free Reading A Impact		Spell Read B Impact		Wilson Reading C Impact		Corrective Reading D Impact	
PSSA Reading	-21.5	#	-23.7		-15.0	#	-0.7	#	-37.3		-33.2	
PSSA Math	-11.8		-29.3		40.8		13.5		-101.7	*	0.2	
Sample Size	178		120		58		49		31		40	

* 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.

REFERENCES

Agronin, M.E., J.M. Holahan, B.A. Shaywitz, and S.E. Shaywitz. "The Multi-Grade Inventory for Teachers." In S.E. Shaywitz and B.A. Shaywitz, eds., *Attention Deficit Disorder Comes of Age*. Austin, TX: PRO-ED, 1992, 29-67.

Hart, B., and T.R. Risley. *Meaningful Differences in the Everyday Experience of Young American Children.* Baltimore: Paul H. Brookes Publishing Co., 1995.

Howe, K.B., and M.M. Shinn. "Standard Reading Assessment Passages for Use in General Outcome Assessment: A Manual Describing Development and Technical Features." Eden Prairie, MN: Edformation, Inc., 2002

Jenkins, J.R., L.S. Fuchs, P. van den Broek, C. Espin, and S.L Deno. "Sources of Individual Differences in Reading Comprehension and Reading Fluency." *Journal of Educational Psychology* 2003, 95: 719-729.

Juel, C. "Learning to Read and Write: A Longitudinal Study of 54 Children from First Through Fourth Grades." *Journal of Educational Psychology*

Wilson, B. *The Wilson Reading System*, Third Edition. Millbury, MA: Wilson Language Training Corp., 2002

Wise, B.W., J. Ring, and R.K. Olson. "Training Phonological Awareness With and Without Explicit

APPENDIX A

DETAILS OF STUDY DESIGN AND IMPLEMENTATION

A. STUDY DESIGN

Within the schools assigned to each of the interventions, we then used student-level randomization to assign students to either the treatment or control group. Figures A.2 through A.5 offer details on this randomization and the student-level data collection for the study.

Torgesen et al. (2006) analyzed principal and teacher survey data, data from the Common Core of Data (CCD), and Pennsylvania state assessment data to compare the schools in the study with other schools in the AIU and schools throughout Pennsylvania and the nation. They found that the study schools had a higher proportion of black students, a lower proportion of Hispanic students, and a lower proportion of students eligible for free or reduced-price lunches than the nation's schools. Compared with other schools in the AIU, the study schools had a higher proportion of black students, but similar proportions of Hispanic students and students eligible for free or reduced-price lunches. Torgesen et al. (2006) also found that the proportion of students scoring at or above the proficient level on the 2003 state fifth-grade reading assessment in the study schools was not significantly different from the proportion for all Pennsylvania schools. Furthermore, they found that across the groups of study schools assigned to the four interventions, the proportions of fifth-graders scoring proficient or higher in 2003 were not significantly different, suggesting that differences in instructional effectiveness that might occur across the interventions examined in this study should not be attributed to differences in effectiveness of the general reading instruction available at these schools.

Figure A-1

Consort Diagram for Schools in Study

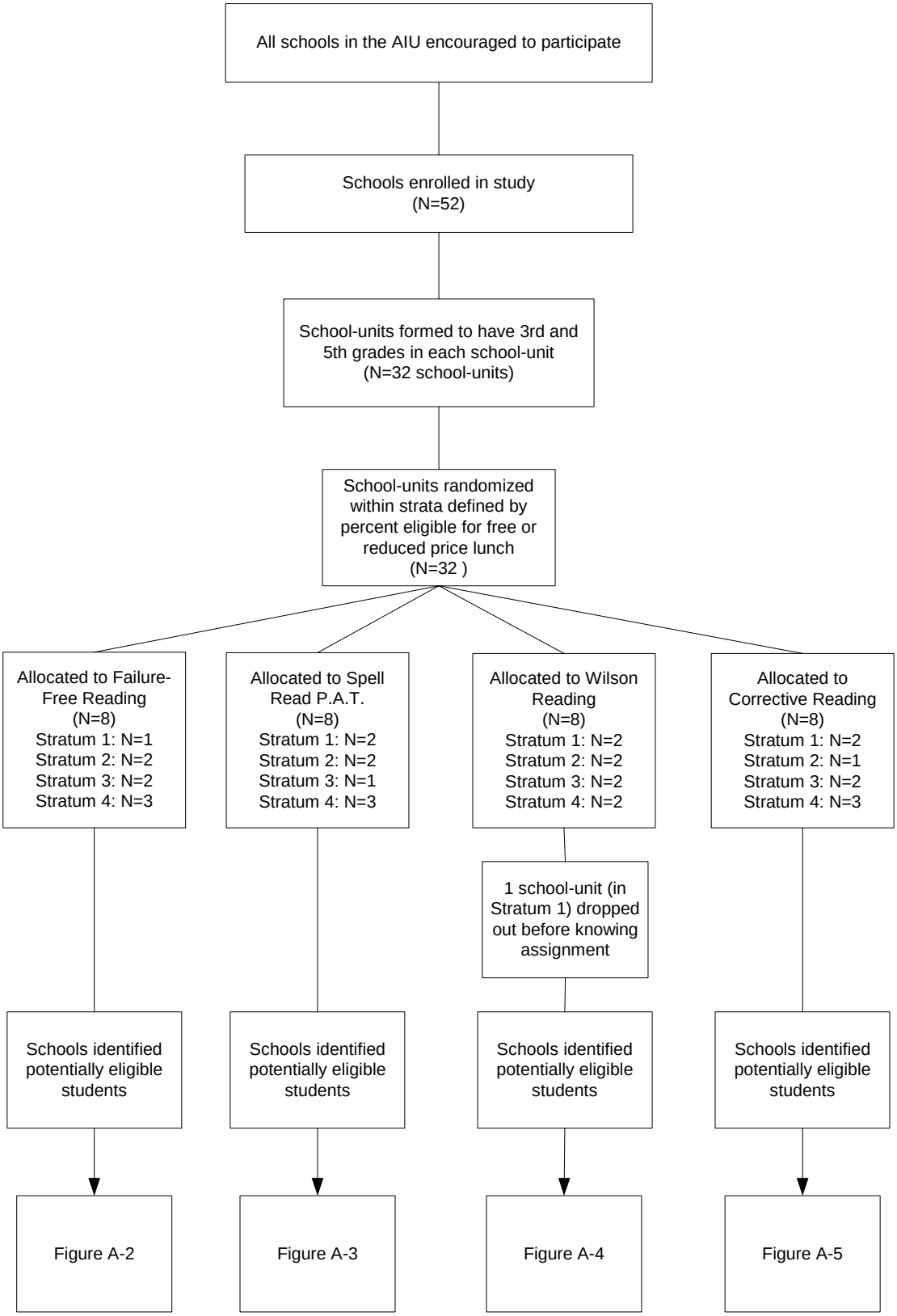
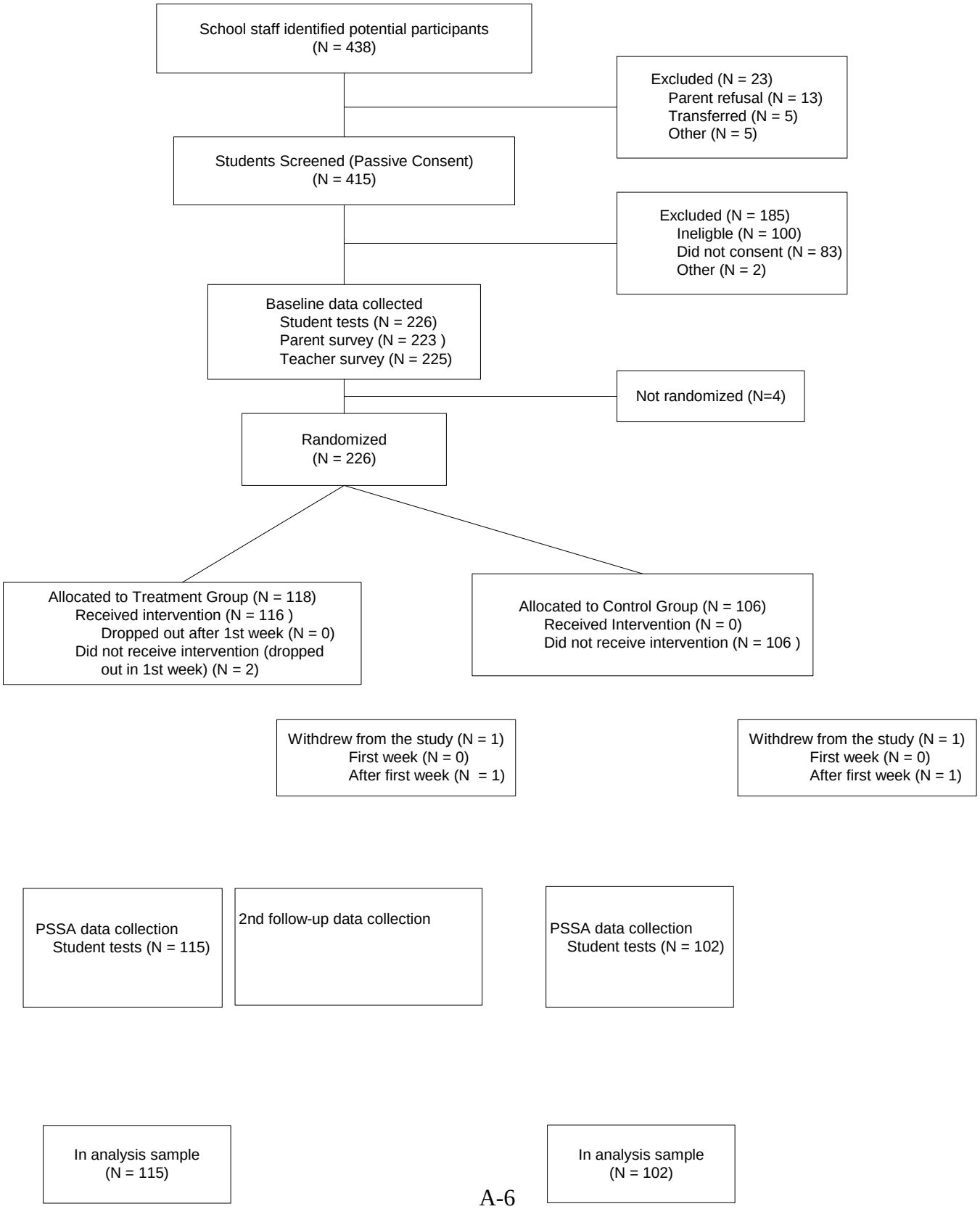


Figure A-2

Progression of Students in Schools Assigned to Failure Free Reading



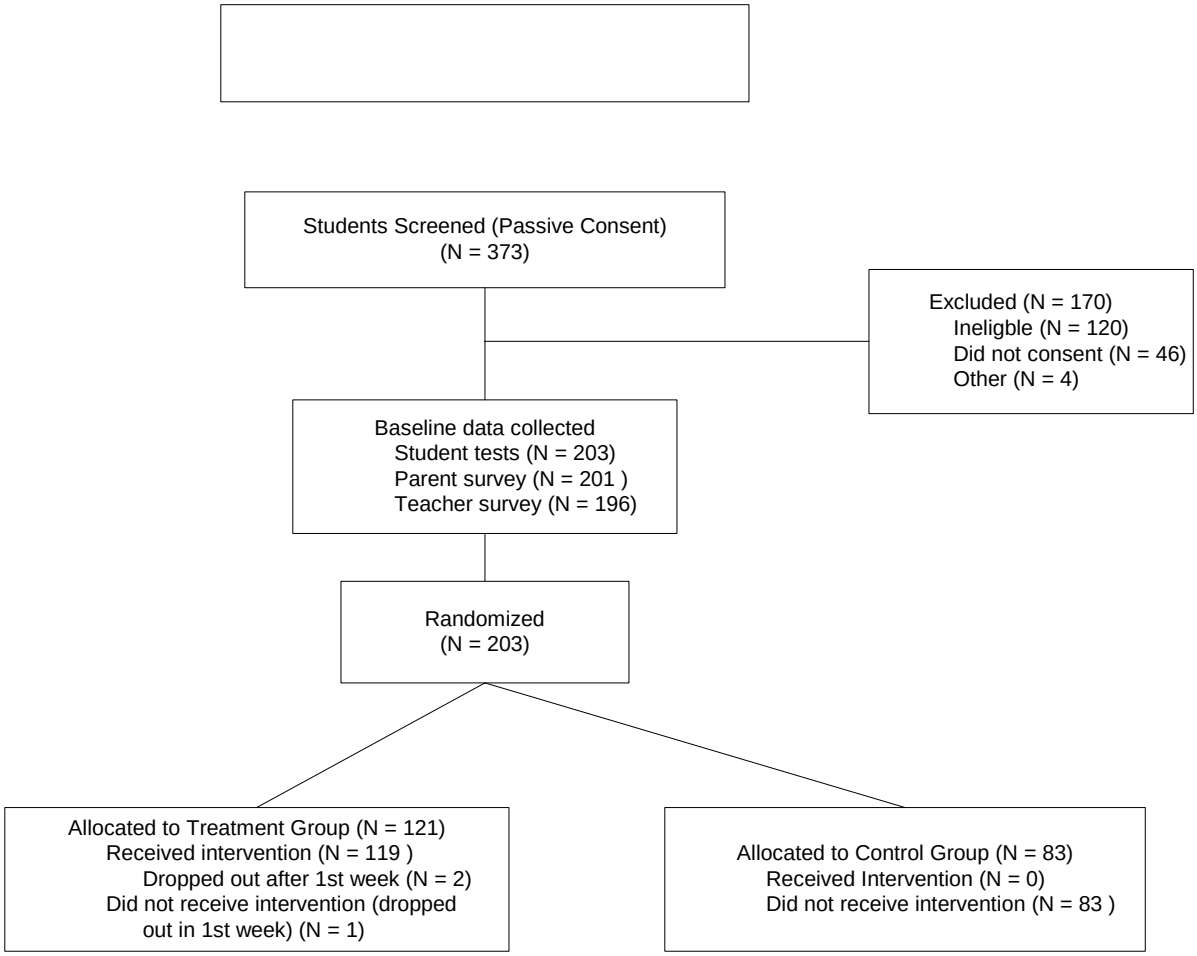

Students Screened (Passive Consent)
(N = 373)
Excluded (N = 170)
Ineligble (N = 120)
Did not consent (N = 46)
Other (N = 4)
Baseline data collected
Student tests (N = 203)
Parent survey (N = 201)
Teacher survey (N = 196)
Randomized
(N = 203)
Allocated to Treatment Group (N = 121)
Received intervention (N = 119)
Dropped out after 1st week (N = 2)
Did not receive intervention (dropped out in 1st week) (N = 1)
Allocated to Control Group (N = 83)
Received Intervention (N = 0)
Did not receive intervention (N = 83)
In analysis sample
(N = 112)
In analysis sample
(N = 82)

Figure A-4

Progression of Students in Schools Assigned to Wilson Reading

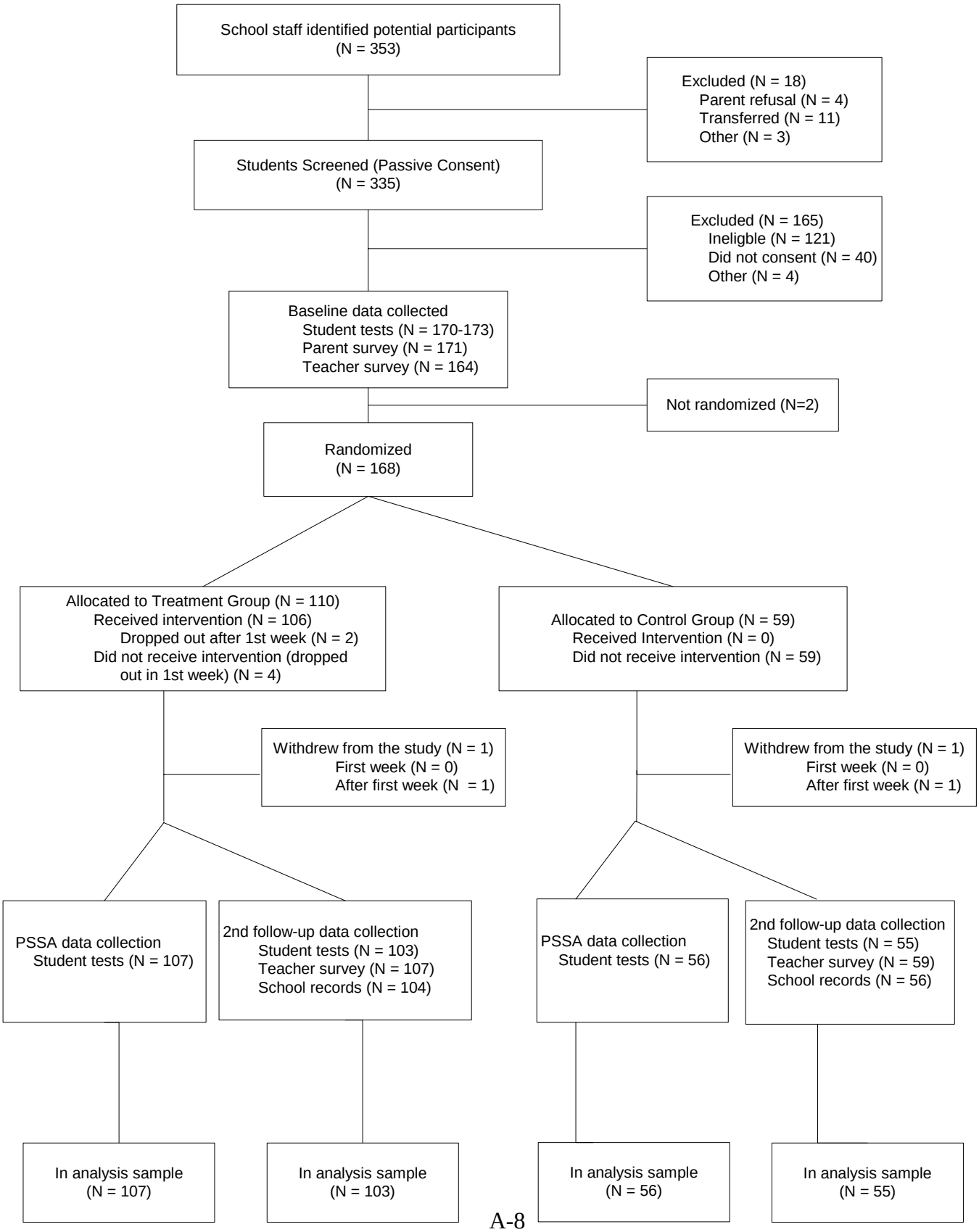
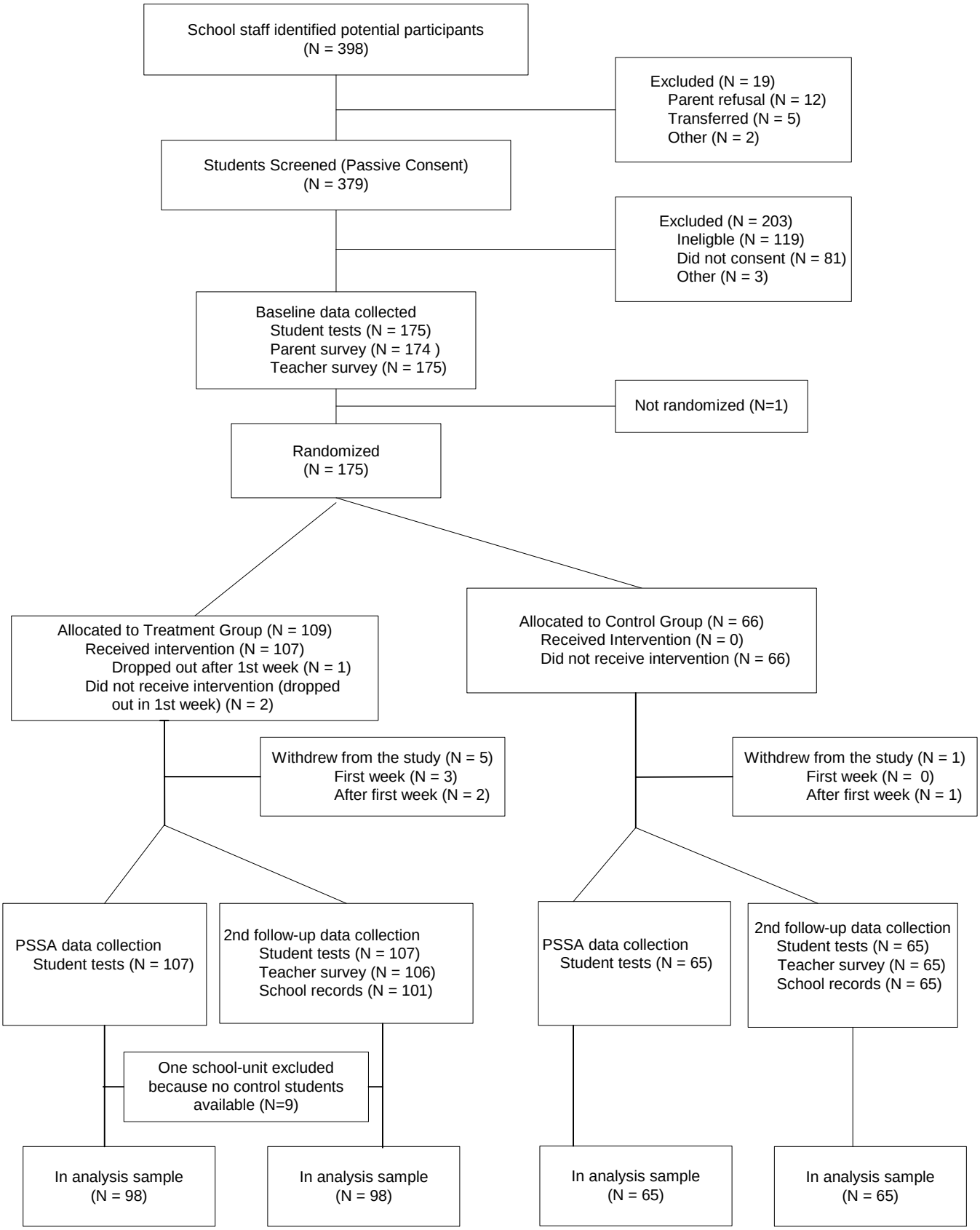


Figure A-5

Progression of Students in Schools Assigned to Corrective Reading



APPENDIX B

DATA COLLECTION

This appendix describes the second follow-up data collection activities, which occurred in the spring semester of the 2004-05 school year. These activities included administering student reading tests, conducting teacher surveys, and extracting school records data. The appendix also describes our approach to gathering Pennsylvania System of School Assessment (PSSA) scores from the 2003-04 school year. Mathematica Policy Research (MPR) staff were responsible for administering the standardized reading tests. Personnel from the Allegheny Intermediate Unit (AIU) worked with MPR staff to coordinate all data collection activities within the AIU, including collecting teacher surveys, school records data, and PSSA scores.[1]

A. ADMINISTERING STANDARDIZED TESTS

B. TEACHER SURVEYS

Near the end of the 2004-05 school year, reading/language arts teachers completed a survey for each

APPENDIX C

WEIGHTING ADJUSTMENTS AND MISSING DATA

A. BASIC WEIGHTING ADJUSTMENTS

Weighting adjustments are utilized to ensure that both the treatment group and the control group reflect

testesponse,d esces, lread5(aiand five percent)5(rfs re evaluae)5sisp sample.

B. WEIGHTING ADJUSTMENTS FOR MISSING DATA

1. Overview of Missing Data Challenges

The preliminary base weights described above were modified slightly to account for missing data on the tests used to assess reading ability and estimate impacts. There are two types of missing data: unit

For each of these strata, we weight up control students in other strata who look like the missing students. This similarity is defined as being in the same blocking stratum and grade, same race, gender, and income (when possible), and similar baseline test scores.[3,4] Because there is no information available on the missing control students, and because at baseline the treatment and control students within a randomization stratum should be only randomly different from one another, we do this by finding control students who look similar to the treatment students in the strata with missing control students.

Priority is given to close matching on the baseline test scores because they should be highly predictive of the outcome test scores. In addition, the matching is generally done so that exact matches on gender, race, and family income categories are obtained when possible—for example, an African American male student in a family with income over $30,000 is matched to other African American male students with family incomes over $30,000 and similar baseline test scores. In this way, control students with the same race, gender, and income level, and similar baseline test scores are weighted up to account for the missing control students in the strata with no control students. The weighting is done inversely proportional to the Mahalanobis distance, so that students with test scores more similar to those of the target are given greater weight than students with test scores less similar.

For example, suppose a stratum with no control students had three treated students, and that the control total for the weighting is six (so that the sum of the weights in the treatment and control groups in that stratum should each equal six). Each treatment student will get a weight of two. In addition, matches will be found for each treatment student in the stratum. Suppose treatment students one and two each have one match, while treatment student three has two matches, with Mahalanobis distances (D) of two and four, respectively. Each of the matches for treatment students one and two will have their weights increased by $w_k = 2 * \frac{1/D}{1/D} = 2$. For treatment student three, the two matches will have their weights increased by $w_1 = 2 * \frac{1/2}{1/2+1/4} = \frac{4}{3}$ and $w_2 = 2 * \frac{1/4}{1/2+1/4} = \frac{2}{3}$, respectively. Note that the sum of these weight increases is $2+2+4/3+2/3=6$, which is the correct control total for that stratum.

3. School Unit With No Control Students

There is one school unit for which there are no control students observed at all, in either third or fifth grade. This case is more complicated than the randomization strata with no control students; it is not possible to estimate school unit parameters in the analysis model when there are no control students observed in an entire school unit. We thus drop this school-unit from the analysis, but weight up other students who look similar to the students in that school unit, to account for their being dropped. This

will preserve the balance achieved through randomization; simply dropping the students from that school unit may degrade such balance.

A weighting adjustment similar to that for the strata with no control students was used here, except that in addition to control matches for the missing control students in this school unit, as described above, we also found treatment matches for the dropped treatment students. The method is similar to the procedure described above, except that the potential matches are restricted to treatment students in the same blocking stratum, grade, and intervention.

4. Nonresponse at Follow-Up

A final source of missing data is student-level (unit) nonresponse. An approach similar to that described above is used to adjust for this nonresponse, where students in the same treatment group, blocking strata, and grade, with the same age, race, family income, and similar baseline test scores, are weighted up to account for the unit nonrespondents. As noted above, we perform separate weight adjustments for our two analysis samples—one sample with the students who took our seven reading tests and one with the students for whom we f 9t-d001de f PSS f Ag test scors.p

APPENDIX D

DETAILS OF STATISTICAL METHODS

A. ESTIMATION OF IMPACTS

The impacts are defined as the regression-adjusted difference in the average achievement scores for the treatment and the control groups. From the hierarchical linear model (HLM), we estimate impacts for each of the four interventions.[5] We also estimate the impact of being assigned to any of the interventions, denoted as the pooled intervention impact (ABCD), as the average of these four intervention impacts.[6]

As discussed in Chapter IV, we can obtain our impact estimates from our model by substituting in the level-two equations (Equations (IV.2) in Chapter IV) into the level-one equation (Equation (IV.1) in Chapter IV) and considering this combined equation for different profiles of students and school units.[7] We assume that each blocking stratum contributes equally to the impacts and construct impacts using an equally weighted average blocking effect. [8]

In addition to estimating impacts for the all students in the third-grade cohort and all students in the fifth-grade cohort, we estimate impacts for subgroups of students within each grade cohort. Being able to estimate impacts for subgroups and to test for differences in impacts among them allows for potentially better targeting of the interventions to, for example, students with especially low phonemic decoding skills. To estimate subgroup impacts, we modify the model specification found in Chapter IV to allow for a general subgroup, S_{ij}, as described below.

---

[5] We used HLM 6 © software published by Scientific Software International, Inc. to obtain the HLM estimates. Parameter estimates are obtained using restricted maximum likelihood (REML), as discussed in Raudenbush and Bryk (2002).

[6] Since we assume that each intervention contributes equally to the pooled intervention impact, we use an equally weighted average of the individual interventions.

[7] For example, the equation pertaining to control students from school units assigned to intervention D is

$$Y_{ij} = \gamma_{00} + \gamma_{20} y^*_{oij} + (1/4)(\xi_{01} + \xi_{02} + \xi_{03}) + (1/4)\ (\xi_{21} + \xi_{22} + \xi_{23}) y^*_{oij} + error$$

And, the equation pertaining to treatments in intervention A is

$$Y_{ij} = \gamma_{00} + \gamma_{01} + \gamma_{10} + \gamma_{11} + (\gamma_{20} + \gamma_{21}) y^*_{oij} + (1/4)(\xi_{01} + \xi_{02} + \xi_{03}) + (1/4)(\xi_{11} + \xi_{12} + \xi_{13}) + (1/4)\ (\xi_{21} + \xi_{22} + \xi_{23}) y^*_{oij} + error.$$

Similar equations can be obtained for each intervention and treatment or control group.

[8] An estimate of impacts for different profiles of students could include those students in a particular intervention within a particular blocking strata. Since we assume blocking strata contribute equally to the impact estimates, we capture the blocking effect by using an average blocking effect, where each blocking strata contributes equally. This average is an equally weighted average of impacts for interventions and combinations of interventions from each of the four blocking strata. For example, impacts for interventions A in blocking strata 1, 2, 3 and 4 are averaged so that each blocking strata contributes equally to impact of intervention A. Thus, for the third-grade cohort, the term $(1/4)(\hat{\xi}_{11} + \hat{\xi}_{12} + \hat{\xi}_{13})$ in Equation (IV.5) captures the blocking effect in intervention A impacts.

B. IMPACTS FOR GENERAL SUBGROUPS MEASURED AT THE INDIVIDUAL LEVEL

The model specification found in Chapter IV is modified as follows to allow for different impacts for a general subgroup, $_{ij}$.

Tables D.1 through D.20 show the significance levels with no adjustment (as presented in Chapter IV), adjusted using the Bonferroni method, and adjusted using the BH method, for the full sample and for each of the subgroups. The odd-numbered tables do the adjustments within tests grouped by type; the even-numbered tables do the adjustments within groups of all seven tests. The corresponding sample

Grade 3 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
TOWRE PDE	*	*	*	*	*	*				*	*	*	*	*	*	*		*
Word Identification	*	*	*	*	*	*							*	*	*	*		
TOWRE SWE	*	*	*	*	*	*	*		*				*	*	*			
AIMSweb							*	*	*	*								
Passage Comprehension	*	*	*				*	*	*									
GRADE																		

Grade 5 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*							*	*	*			
TOWRE PDE				*	*	*												
Word Identificatio																		

Table D.2

3rd and 5th Grade Cohorts: Comparison of Significance Levels, All Tests Grouped Together

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 3 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*		
TOWRE PDE	*	*	*	*	*	*				*	*	*	*	*	*	*		
Word Identification	*	*	*	*	*	*							*	*	*	*		
TOWRE SWE	*	*	*	*		*	*		*				*		*			
AIMSweb							*	*	*	*		*						
Passage Comprehension	*		*				*	*	*									
GRADE																		

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 5 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*							*	*	*			
TOWRE PDE				*														
Word Identification																		
TOWRE SWE										*	*	*						
AIMSweb	*			*														
Passage Comprehension																		
GRADE																		

* Impact is statistically significant at the 0.05 level.

Note: "Unadj." is unadjusted p-values, "Bonf." referes to p-values adjusted using Bonferroni procedure, "BH" refers to p-values adjusted using Benjamini and Hochberg (1995) procedure.

Table D.4

3rd and 5th Grade Cohorts with Low Word Attack Scores: Comparison of Significance Levels, All Tests Grouped Together

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 3 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*						*								
TOWRE PDE	*		*	*						*								
Word Identification	*	*	*	*									*	*	*			
TOWRE SWE																		
AIMSweb							*			*								
Passage Comprehension																		
GRADE																		

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 5 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*				*	*	*	*		*	*		
TOWRE PDE				*														
Word Identification																		
TOWRE SWE	*						*			*								
AIMSweb	*			*									*					
Passage Comprehension																		
GRADE	*			*		*							*	*	*			

* Impact is statistically significant at the 0.05 level.

Note: "Unadj." is unadjusted p-values, "Bonf." referes to p-values adjusted using Bonferroni procedure, "BH" refers to p-values adjusted using Benjamini and Hochberg (1995) procedure.

Table D.5

3rd and 5th Grade Cohorts with High Word Attack Scores: Comparison of Significance Levels, Tests Grouped by Type

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 3 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*	*						*		*	*	*	*
TOWRE PDE	*	*	*	*	*	*				*	*	*	*	*	*	*	*	*
Word Identification	*	*	*	*	*	*	*		*				*	*	*			
TOWRE SWE	*	*	*	*	*	*	*	*	*				*	*	*	*	*	*
AIMSweb	*		*	*		*							*		*			
Passage Comprehension	*	*	*				*	*	*									
GRADE							*		*				*	*	*			

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 5 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack													*	*	*			
TOWRE PDE																		
Word Identification	*																	
TOWRE SWE										*	*	*						
AIMSweb																		
Passage Comprehension	*	*	*													*		*
GRADE							*	*	*							*		*

* Impact is statistically significant at the 0.05 level.

Note: "Unadj." is unadjusted p-values, "Bonf." referes to p-values adjusted using Bonferroni procedure, "BH" refers to p-values adjusted using Benjamini and Hochberg (1995) procedure.

Grade 3 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*	*	*	*				*	*	*	*		*
TOWRE PDE	*	*	*	*	*	*							*	*	*	*	*	*
Word Identification													*	*	*			
TOWRE SWE																		
AIMSweb																		
Passage Comprehension																		
GRADE																		

Grade 5 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*							*	*	*			
TOWRE PDE	*	*	*	*	*	*												
Word Identification																		
TOWRE SWE	*	*	*	*	*	*				*	*	*				*		
AIMSweb																		
Passage Comprehension				*		*							*	*	*			
GRADE	*	*	*	*		*	*						*	*	*	*		

* Impact is statistically significant at the 0.05 level.
Note: "Unadj." is unadjusted p-values, "Bonf." referes to p-values adjusted using Bonferroni procedure, "BH" refers to p-values adjusted using Benjamini765(*)947(**)7nd[(*)13(I)-8(m)16(pa73(amun)7(g)7( )-8(Beod5d)-7(ued u)7(s)4 Tclue

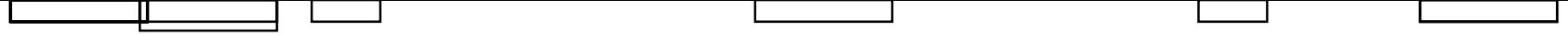


Table D.10

3rd and 5th Grade Cohorts with High Screening Peabody Picture Vocabulary Test Scores: Comparison of Significance Levels, All Tests Grouped Together

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 3 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*				*	*	*				*		
TOWRE PDE	*	*	*	*	*	*				*	*	*						
Word Identification	*	*	*	*	*	*	*		*				*	*	*			
TOWRE SWE	*		*				*		*									
AIMSweb	*	*	*	*	*	*	*	*	*	*			*	*	*			
Passage Comprehension	*		*				*	*	*									
GRADE																		

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 5 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*			*		*				*	*	*						
TOWRE PDE	*			*	*	*				*	*	*						
Word Identification																		
TOWRE SWE										*								
AIMSweb	*	*	*	*	*	*										*	*	*
Passage Comprehension																		
GRADE																		

* Impact is statistically significant at the 0.05 level.

Note: "Unadj." is unadjusted p-values, "Bonf." referes to p-values adjusted using Bonferroni procedure, "BH" refers to p-values adjusted using Benjamini and Hochberg (1995) procedure.

Table D.11

3rd and 5th Grade Cohorts with Low Baseline Word Attack Scores and Low Screening Peabody Picture Vocabulary Test Scores: Comparison of Significance Levels, Tests Grouped by Type

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 3 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack TOWRE PDE	*	*	*	*	*	*										*		
Word Identification TOWRE SWE AIMSweb	*			*	*	*							*	*	*			
Passage Comprehension GRADE																		

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 5 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*				*	*	*				*	*	*
TOWRE PDE	*	*	*	*	*	*												
Word Identification TOWRE SWE	*									*								
AIMSweb	*			*														
Passage Comprehension GRADE	*	*	*	*	*	*				*	*	*				*		

* Impact is statistically significant at the 0.05 level.

Note: "Unadj." is unadjusted p-values, "Bonf." referes to p-values adjusted using Bonferroni procedure, "BH" refers to p-values adjusted using Benjamini and Hochberg (1995) procedure.

Table D.13

3rd and 5th Grade Cohorts with Low Baseline Word Attack Scores and High Screening Peabody Picture Vocabulary Test Scores: Comparison of Significance Levels, Tests Grouped by Type

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 3 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack																*	*	*
TOWRE PDE										*	*	*						
Word Identification	*		*										*	*	*			
TOWRE SWE																		
AIMSweb	*	*	*	*			*											
Passage Comprehension	*						*	*	*									
GRADE							*	*	*									

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 5 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*				*	*	*	*	*	*			
TOWRE PDE	*	*	*	*	*	*				*	*	*	*		*	*		
Word Identification							*											
TOWRE SWE	*	*	*	*	*	*				*	*	*						
AIMSweb													*					
Passage Comprehension																		
GRADE	*	*	*	*	*	*							*	*	*			

* Impact is statistically significant at the 0.05 level.

Note: "Unadj." is unadjusted p-values, "Bonf." referes to p-values adjusted using Bonferroni procedure, "BH" refers to p-values adjusted using Benjamini and Hochberg (1995) procedure.

Table D.14

3rd and 5th Grade Cohorts with Low Baseline Word Attack Scores and High Screening Peabody Picture Vocabulary Test Scores: Comparison of Significance Levels, All Tests Grouped Together

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 3 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack																*		
TOWRE PDE										*	*	*						
Word Identification	*												*	*	*			
TOWRE SWE																		
AIMSweb	*			*			*											
Passage Comprehension	*						*	*	*									
GRADE							*											

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 5 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*				*		*	*	*	*			
TOWRE PDE	*	*	*	*	*	*				*	*	*	*		*	*		
Word Identification							*											
TOWRE SWE	*		*	*		*				*	*	*						
AIMSweb													*		*			
Passage Comprehension																		
GRADE	*		*	*	*	*							*		*			

* Impact is statistically significant at the 0.05 level.

Note: "Unadj." is unadjusted p-values, "Bonf." referes to p-values adjusted using Bonferroni procedure, "BH" refers to p-values adjusted using Benjamini and Hochberg (1995) procedure.

Table D.15

3rd and 5th Grade Cohorts with High Baseline Word Attack Scores and High Screening Peabody Picture Vocabulary Test Scores: Comparison of Significance Levels, Tests Grouped by Type

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 3 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*	*			*	*	*	*		*	*	*	*
TOWRE PDE	*	*	*	*	*	*				*	*	*	*	*	*			
Word Identification	*			*	*	*										*	*	*
TOWRE SWE	*	*	*	*	*	*	*	*	*				*	*	*	*	*	*
AIMSweb													*	*	*			
Passage Comprehension	*	*	*				*	*	*									
GRADE																		

	All Interventions			Word-level Interventions														
Grade 5 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack																		
TOWRE PDE	*			*														
Word Identification										*								
TOWRE SWE																		
AIMSweb																*	*	*
Passage Comprehension																		
GRADE	*	*	*	*			*	*	*							*	*	*

* Impact is statistically significant at the 0.05 level.

Note: "Unadj." is unadjusted p-values, "Bonf." referes to p-values adjusted using Bonferroni procedure, "BH" refers to p-values adjusted using Benjamini and Hochberg (1995) procedure.

Table D.16

3rd and 5th Grade Cohorts with High Baseline Word Attack Scores and High Screening Peabody Picture Vocabulary Test Scores: Comparison of Significance Levels, All Tests Grouped Together

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 3 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*	*			*	*	*	*			*	*	*
TOWRE PDE	*	*	*	*	*	*				*	*	*	*		*			
Word Identification	*			*		*										*	*	*
TOWRE SWE	*	*	*	*	*	*	*	*	*				*		*	*	*	*
AIMSweb													*	*	*			
Passage Comprehension	*	*	*				*	*	*									
GRADE																		

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 5 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack																		
TOWRE PDE	*			*														
Word Identification										*								
TOWRE SWE																		
AIMSweb																*	*	*
Passage Comprehension																		
GRADE	*			*			*									*	*	*

* Impact is statistically significant at the 0.05 level.

Note: "Unadj." is unadjusted p-values, "Bonf." referes to p-values adjusted using Bonferroni procedure, "BH" refers to p-values adjusted using Benjamini and Hochberg (1995) procedure.

Table D.17

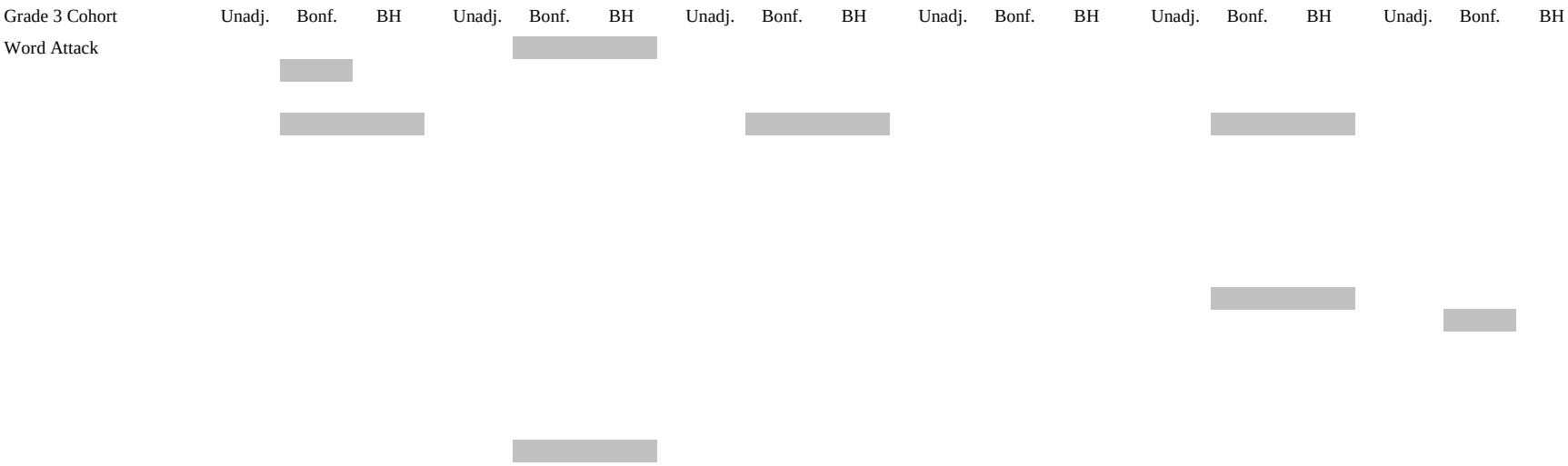
3rd and 5trade Corts Eligible for Free or Reduced-Price Scol Lunc Comparison of Significance Levels, Tests Grouped -4iby Type

Grade 3 Cohort	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*		*	*	*	*				*	*	*			
TOWRE PDE	*	*	*	*	*	*				*	*	*	*	*	*			
Word Identification													*	*	*			
TOWRE SWE	*						*						*		*			
AIMSweb																		
Passage Comprehension																		
GRADE																		

Grade 5 Cohort	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*							*	*	*			
TOWRE PDE																*	*	*
Word Identification																		
TOWRE SWE	*	*	*	*	*	*				*	*	*				*	*	*
AIMSweb																		
Passage Comprehension																		
GRADE	*	*	*	*			*	*	*	*	*	*						

* Impact is statistically significant at the 0.05 level.

Note: "Unadj." is unadjusted p-values, "Bonf." referes to p-values adjusted using Bonferroni procedure, "BH" refers to p-values adjusted using Benjamini and Hocg (1995) procedure.


Grade 3 Cohort
Word Attack
Unadj.
Bonf.
BH
Unadj.
Bonf.
BH
Unadj.
Bonf.
BH
Unadj.
Bonf.
BH
Unadj.
Bonf.
BH
Unadj.
Bonf.
BH

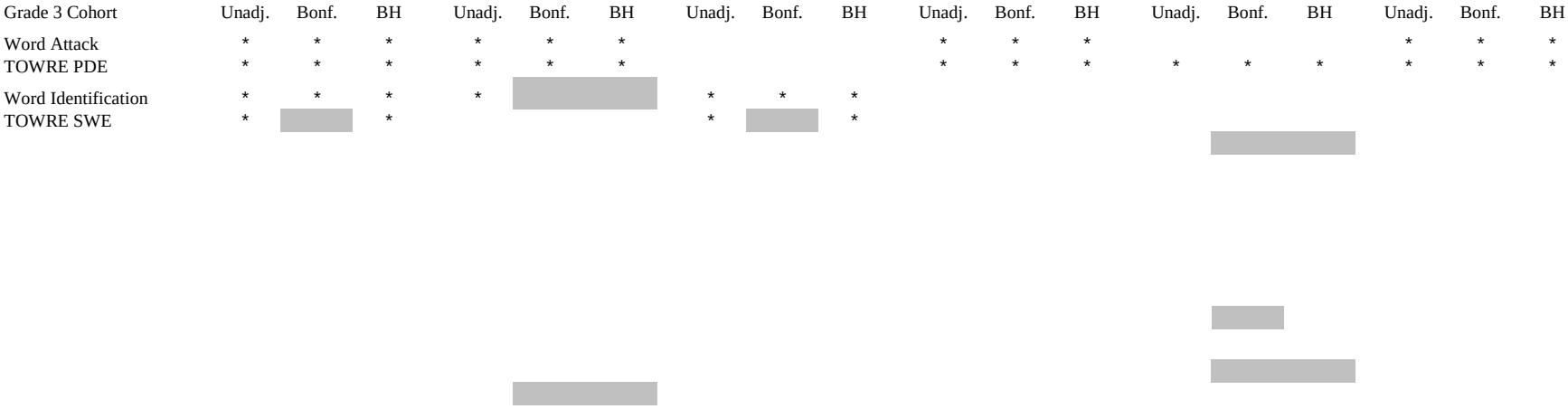

Grade 3 Cohort
Unadj.
Bonf.
BH
Unadj.
Bonf.
BH
Unadj.
Bonf.
BH
Unadj.
Bonf.
BH
Unadj.
Bonf.
BH
Unadj.
Bonf.
BH
Word Attack
TOWRE PDE
Word Identification
TOWRE SWE

Table D.20

3rd and 5th Grade Cohorts Not Eligible for Free or Reduced-Price School Lunch: Comparison of Significance Levels, All Tests Grouped Together

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 3 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*		*	*		*				*	*	*				*		*
TOWRE PDE	*	*	*	*	*	*				*	*	*	*	*	*	*	*	*
Word Identification	*	*	*	*			*		*									
TOWRE SWE	*						*											
AIMSweb	*	*	*	*		*	*	*	*				*					
Passage Comprehension	*		*				*	*	*									
GRADE																		

	All Interventions			Word-level Interventions			Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
Grade 5 Cohort	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH	Unadj.	Bonf.	BH
Word Attack	*	*	*	*	*	*				*	*	*	*	*	*			
TOWRE PDE	*			*	*	*				*	*	*	*					
Word Identification																		
TOWRE SWE													*					
AIMSweb				*						*		*				*	*	*
Passage Comprehension																		
GRADE																		

* Impact is statistically significant at the 0.05 level.

Note: "Unadj." is unadjusted p-values, "Bonf." referes to p-values adjusted using Bonferroni procedure, "BH" refers to p-values adjusted using Benjamini and Hochberg (1995) procedure.

APPENDIX E

INTERVENTION IMPACTS ON SPELLING AND CALCULATION

Table E.1

Impacts for 3rd and 5thl4c6 1e8 Tc -1 Td[B)-7((2)5(f)8(o)6(r)-7(3r) Td6J0.07Td6Jr(2)5s238.2 273.06 Id[Im Td[A)11(BCD)]TJ11.515 0 Td

		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 Cohort	Baseline	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact
WJ-III Spelling	88.5	2.4	1.3 *	2.9	0.9	0.8	2.6	2.3	0.9	3.1	0.9	3.3	0.8
WJ-III Calculation	95.4	1.9	4.3 *	2.4	3.6	0.6	6.6 *	5.3	0.4	-4.0	7.6	5.8	2.7
Sample Size	328												

		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 5 Cohort	Baseline	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact
WJ-III Spelling	90.9	1.6	-0.5 #	1.4	-0.2	2.2	-1.5 * #	1.5	1.2	0.3	-0.2	2.5	-1.4
WJ-III Calculation	94.2	5.4	-1.5 * #	4.4	-0.8	8.3	-3.7 * #	2.2	0.7	4.8	-2.9 * #	6.2	-0.2
Sample Size	399												

* Impact is statistically significant at the 0.05 level.
# Impact is significantly different than the impact for the 3rd grade cohort at the 0.05 level.

.3

Impacts for 3rd and 5th Grade Cohorts with High Baseline Word Attack Scores

All Interventions					
ABCD	BCD			C	D
Impact	Impact	Impact	Impact	Impact	

le E.4

Impacts for 3rd and 5th Grade Cohorts with Low Screening Peabody Picture Vocabulary Test Scores

		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 Cohort	Baseline	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact
WJ-III Spelling	87.8	3.9	-1.1	4.7	-1.6	1.5	0.4	5.9	-1.5	5.4	-1.7	2.7	-1.4 *
WJ-III Calculation	93.3	6.0	-0.5	6.8	-1.3	3.3	1.9	5.7	1.2	6.1	-2.9	8.7	-2.2 * #
Sample Size	146												

		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 5 Cohort	Baseline	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact
WJ-III Spelling	89.2	2.1	-0.7	1.7	-0.1	3.3	-2.6 *	3.0	0.7	2.1	-1.1	0.2	0.2
WJ-III Calculation	92.2	3.3	0.5	2.0	1.4	7.0	-2.2	-0.6	4.0 * #	1.9	0.0	4.9	0.3
Sample Size	194												

le E.5

Impacts for 3rd and 5th Grade Cohorts with High

Table E.6

Impacts for 3rd and 5th Grade Cohorts with Low Baseline Word Attack and Low Screening Peabody Picture Vocabulary Test Scores

		All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
Grade 3 Cohort	Baseline	Control Gain	ABCD Impact	Control Gain	BCD Impact	Control Gain	A Impact	Control Gain	B Impact	Control Gain	C Impact	Control Gain	D Impact
WJ-III Spelling	83.0	6.2	-2.4 * #	6.7	-2.9 * #	4.6	-1.1	8.0	-2.4	8.2	-1.2	3.9	-5.1
WJ-III Calculation	89.6	3.5	2.8	3.3	4.2	4.1	-1.4 #	10.5	-4.4	-2.3	10.9	1.9	6.3
Sample Size	80												
		All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
Grade 5 Cohort	Baseline	Control Gain	ABCD Impact	Control Gain	BCD Impact	Control Gain	A Impact	Control Gain	B Impact	Control Gain	C Impact	Control Gain	D Impact
WJ-III Spelling	84.3	2.7	-2.0 *	2.7	-2.0	2.6	-2.2	4.1	-2.2 #	2.8	-3.1	1.2	-0.5 *
WJ-III Calculation	89.9	4.7	-1.0	4.2	0.0	5.9	-3.9	-2.9	5.4 * #	7.3	-3.6	8.3	-1.9
Sample Size	106												

* 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 E.8

Grade 3 Cohort	Baseline	Control Gain	ABCD Impact	Control Gain	BCD Impact	Control Gain		Control Gain		Control Gain		Control Gain	
WJ-III Spelling	94.8	3.3	-0.1	3.4	-1.0	2.8	2.5	-0.9	1.7	5.9	-2.7	5.3	-1.9
WJ-III Calculation	100.0	3.0	2.7	4.5	-0.6 #	-1.6	12.7 * #	6.2	-4.5 #	-0.6	3.8	7.9	-1.0
Sample Size	93												

Corrective Reading

Grade 5 Cohort	Baseline	Control Gain		Control Gain		Control Gain		Control Gain		Control Gain		Control Gain		
WJ-III Spelling	96.0	2.7	-2.5 * #	2.8	-2.7 * #	2.4	-1.9	0.6	0.0	2.6	-2.1	5.2	-5.9 * #
WJ-III Calculation	96.8	5.9	-2.2	5.6	-2.6	6.8	-1.0	6.9	-7.3 #	4.6	-1.3	5.2	0.8
Sample Size	117												

* 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 E.10

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

		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 Cohort	Baseline	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact
WJ-III Spelling	89.4	1.7	2.3	1.0	2.9	3.9	0.2	1.3	2.5 *	-1.9	5.5	3.5	0.9
WJ-III Calculation	95.8	-3.3	9.0	-4.4	9.4	0.0	7.8	5.5	-0.7	-22.1	25.5	3.5	3.5
Sample Size	139												

		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 5 Cohort	Baseline	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact	Gain	Impact
WJ-III Spelling	92.7	2.2	-1.5 *	2.0	-1.5	2.8	-1.5	-0.2	3.3 * #	1.0	-3.5	5.3	-4.3 *
WJ-III Calculation	95.8	5.8	-1.1	5.4	-0.4	6.8	-3.2	3.4	1.3	7.2	-4.9 *	5.7	2.2 #
Sample Size	180												

* 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 E.11

Test Statistics and P-values of Impacts for 3rd and 5th Grade Cohorts

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	3.85	0.05	1.87	0.17	2.96	0.08	0.46	0.50	1.12	0.29	0.93	0.50
WJ-III Calculation	5.34	0.02	2.58	0.10	4.51	0.03	0.02	0.50	2.19	0.13	0.75	0.50

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	0.60	0.50	0.04	0.50	4.04	0.04	0.80	0.50	0.05	0.50	0.92	0.50
WJ-III Calculation	4.02	0.04	0.98	0.50	4.58	0.03	0.23	0.50	9.92	0.00	0.02	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	0.31	0.50	0.16	0.50	1.82	0.17	0.03	0.50	0.02	0.50	0.31	0.50
WJ-III Calculation	4.42	0.03	4.67	0.03	0.01	0.50	0.21	0.50	3.04	0.08	1.48	0.22

Grade 5 Cohort	Test	*p*	Test	*p*	TeTestpTest	

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	7.01	0.01	2.34	0.12	7.46	0.01	0.41	0.50	1.68	0.19	1.12	0.29
WJ-III Calculation	0.06	0.50	4.27	0.04	21.76	0.00	0.14	0.50	2.93	0.08	5.07	0.02

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	3.89	0.05	2.11	0.14	2.84	0.09	1.20	0.27	1.17	0.28	11.42	0.00
WJ-III Calculation	1.05	0.31	1.34	0.25	0.01	0.50	0.24	0.50	3.01	0.08	0.69	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p*

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	5.22	0.02	1.50	0.22	5.77	0.02	0.39	0.50	0.08	0.50	1.23	0.27
WJ-III Calculation	3.84	0.05	2.19	0.13	2							

Table E.16

Test Statistics and P-values of Impacts for 3rd and 5th Grade Cohorts with Low Baseline Word Attack and Low Screening Peabody Picture Vocabulary Test Scores

	All Interventions ABCD		Word-level Interventions BCD		Failure Free Reading A		Spell Read B		Wilson Reading C		Corrective Reading D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	3.81	0.05	6.00	0.01	0.07	0.50	1.35	0.24	0.23	0.50	1.98	0.16
WJ-III Calculation	0.63	0.50	0.92	0.50	0.08	0.50	0.65	0.50	1.39	0.24	0.53	0.50

	All Interventions ABCD		Word-level Interventions BCD		Failure Free Reading A		Spell Read B		Wilson Reading C		Corrective Reading D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	4.51	0.03	2.10	0.14	1.62	0.20	2.78	0.09	2.58	0.10	0.04	0.50
WJ-III Calculation	0.30	0.50	0.00	0.50	2.00	0.15	9.29	0.00	0.94	0.50	0.14	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table E.17

Test Statistics and P-values of Impacts for 3rd and 5th Grade Cohorts With Low Baseline Word Attack and High Peabody Picture Vocabulary Test Scores

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	1.83	0.17	0.71	0.50	4.36	0.03	0.05	0.50	0.05	0.50	1.38	0.24
WJ-III Calculation	1.35	0.24	2.18	0.14	0.11	0.50	1.64	0.20	2.15	0.14	0.00	0.50

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	2.23	0.13	3.47	0.06	0.38	0.50	0.99	0.50	0.02	0.50	5.02	0.02
WJ-III Calculation	3.63	0.05	0.17	0.50	13.36	0.00	0.08	0.50	1.64	0.20	0.06	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	0.02	0.50	1.11	0.29	2.16	0.14	1.70	0.19	1.44	0.23	1.89	0.17
WJ-III Calculation	1.55	0.21	0.05	0.50	13.18	0.00	2.26	0.13	0.57	0.50	0.09	0.50

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	7.96	0.01	6.05	0.01	1.32	0.25	0.00	0.50	0.73	0.50	13.86	0.00
WJ-III Calculation	1.64	0.20	1.64	0.20	0.10	0.50	2.41	0.12	0.27	0.50	0.30	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear (Sq)5254(s)TJEpq th Bn7est" refey2 B-1950se195we195w d

Grade 3 Cohor

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*

Table E.21

Tests for Differences in Impacts for 3rd and 5th Grade Cohorts

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	3.84	0.05	1.05	0.31	5.87	0.02	0.01	0.91	0.70	0.40	1.73	0.19
WJ-III Calculation	8.37	0.00	3.42	0.06	8.39	0.00	0.01	0.92	4.02	0.05	0.70	0.40

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software.

Table E.23

Tests for Differences in Impacts for 3rd and 5th Grade Cohorts with Low Screening Peabody Picture Vocabulary Test Scores

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	3.56	0.06	2.27	0.13	0.81	0.50	1.19	0.27	0.44	0.50	0.74	0.50
WJ-III Calculation	3.42	0.06	3.25	0.07	1.43	0.23	0.10	0.50	6.62	0.01	1.80	0.18

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	0.06	0.50	0.37	0.50	0.94	0.50	0.00	0.50	0.09	0.50	0.94	0.50
WJ-III Calculation	2.59	0.10	2.74	0.09	0.40	0.50	5.61	0.02	1.20	0.27	0.03	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling												

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	1.85	0.17	2.82	0.09	0.00	0.50	0.81	0.50	3.44	0.06	1.99	0.15
WJ-III Calculation	0.84	0.50	4.04	0.04	6.62	0.01	4.62	0.03	0.33	0.50	1.80	0.18

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	11.73	0.00	13.29	0.00	0.03	0.50	0.55	0.50	2.47	0.11	10.30	0.00
WJ-III Calculation	0.19	0.50	1.06	0.31	0.98	0.50	4.61	0.03	0.03	0.50	0.30	0.50

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	0.01	0.50	0.48	0.50	3.19	0.07	1.00	0.32	0.57	0.50	0.20	0.50
WJ-III Calculation	1.38	0.24	1.25	0.26	0.15	0.50	1.58	0.21	2.15	0.14	0.02	0.50

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
WJ-III Spelling	0.00	0.50	0.02	0.50	0.02	0.50	4.60	0.03	1.17	0.28	2.07	0.15
WJ-III Calculation	2.40	0.12	1.65	0.20	1.23	0.27	1.84	0.17	0.41	0.50	4.60	0.03

APPENDIX F

INSTRUCTIONAL GROUP CLUSTERING

For treatment students, the interventions were administered in instructional groups of three students

cannot be statistically distinguished from zero. For example, on the Passage Comprehension test for the third grade cohort, the word-level intervention impact (BCD) is measured with low precision in both the two-level model and the three-level model, thus making the small ratio of variances practically unimportant. Overall, there is generally a gain in precision when using the three-level model; however, the differences do not change the substantive conclusions, and would change the significance level of an impact only in a case with borderline significance. We therefore base our conclusions on impact estimates derived from the two-level model.

Grade 3 Cohort

Word Attack	4.58 *	4.85 *	3.79 *	5.07 *	5.17 *	4.30 *
TOWRE PDE	3.46 *	4.30 *	0.95	5.10 *	3.90 *	3.90
Word Identification	2.17 *	2.47 *	1.26	1.15	4.07 *	2.18
TOWRE SWE	2.19 *	2.30 *	1.86 *	0.82	3.19 *	2.88
AIMSweb	6.30 *	6.22	6.53 *	6.70 *	4.29	7.68
Passage Comprehension	1.95 *	1.35	3.75 *	-0.52	2.99	1.59
GRADE	0.79	0.07	2.95	1.30	-0.48	-0.61

Grade 3 Cohort

Word Attack	2.39 *	3.49 *	-0.90	3.53 *	6.50 *	0.44
TOWRE PDE	1.83 *	2.38 *	0.17	3.47 *	1.45	2.21
Word Identification	-0.75	-0.79	-0.62	-1.27	0.16	-1.27
TOWRE SWE	1.71 *	1.68	1.78	3.80 *	0.59	0.65
AIMSweb	-2.85	-2.66	-3.42	-2.69	-1.61	-3.67
Passage Comprehension	-1.06	-0.76	-1.97	-0.77	0.26	-1.77
GRADE	0.51	0.83	-0.45	-0.3Td.42 94i358(*)-4850.83		

APPENDIX G

BASELINE CHARACTERISTICS OF THE PSSA SAMPLE

Table G.1

Baseline Characteristics of the Four Intervention Groups and the Control Group,
Analysis Sample: 3rd Grade (PSSA Sample)

Baseline Means	Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading	
	Treat.	Cont.		Treat.	Cont.		Treat.	Cont.		Treat.	Cont.
Student Characteristics											
Age	8.5	8.7		8.8	8.7		8.7	8.6		8.7	8.6
Male (%)	53	58		73	59		39	21		56	44
Hispanic (%)	a	a		a	a		a	a		a	a
Race--White (%)	77	83		65	68		55	68		74	82
Race--African American (%)	23	17		35	32		45	32		26	18
Race--Other (%)	a	a		a	a		a	a		a	a
Family income less than $30,000 (%)	43	39		57	49		48	63		41	60
Family income between $30,000 and $60,000 (%)	46	45		20	24		32	37		41	26
Family income over $60,000 (%)	11	15		23	27		a	a	*	18	14
Eligible for free or reduced price lunch (%)	47	47		46	36		36	70	*	42	49
Has any learning or other disability (%)	40	48		35	25		34	19		30	32
Mother has bachelor's degree or higher (%)	14	9	*	13	15		a	a		19	11
Screening Tests											
TOWRE Sight Word Efficiency	84.2	81.9		85.4	84.7		86.8	83.9		85.6	83.7
TOWRE Phonemic Decoding Efficiency	84.4	85.4		85.7	85.0		86.1	85.8		85.9	87.5
Peabody Picture Vocabulary Test--Revised	93.9	95.2		95.5	97.8		90.4	89.8		97.4	99.9
Baseline Tests											
WRM Word Identification	88.9	87.4		89.5	87.2	*	90.6	89.2		89.7	89.1
TOWRE Phonemic Decoding Efficiency	84.5	84.2		86.2	84.6		87.0	86.3		87.3	87.0
WRM Word Attack	90.2	89.1		93.8	91.4		94.7	94.2		93.9	95.8
TOWRE Sight Word Efficiency	87.3	84.6		89.3	86.6	*	89.0	83.3	*	87.0	86.5
AIMSWeb (Raw score)	38.4	33.5		46.8	41.4		49.3	38.0	*	43.5	41.5
WRM Passage Comprehension	90.9	88.9		95.2	89.9	*	93.8	92.0		94.2	92.9
GRADE	86.4	85.0		87.8	83.9		88.6	85.1		89.7	88.2
Woodcock Johnson Spelling	90.0	86.8		89.4	89.0		89.3	86.7		90.3	87.5
Woodcock Johnson Calculation	93.2	97.1	*	99.3	95.1	*	96.9	91.1		96.8	98.3
Sample Size	53	36		56	36		53	18		43	34

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

Note: All test scores are shown as standard scores, unless otherwise indicated. 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.

* Difference between treatment and control groups is statistically significant at the 0.05 level.

Difference across the four interventions (with treatment and control groups pooled within each intervention) is statistically significant at the 0.05 level.

a Values suppressed to protect student confidentiality.

Table G.2

Baseline Characteristics of the Four Intervention Groups and the Control Group,
Analysis Sample: 5th Grade (PSSA Sample)

Baseline Means	Failure Free Reading			Spell Read			Wilson Reading			Corrective Reading		
	Treat.	Cont.		Treat.	Cont.		Treat.	Cont.		Treat.	Cont.	
Student Characteristics												
Age	10.6	10.6		10.9	10.7		10.8	10.5	*	10.7	10.6	
Male (%)	53	51		55	58		54	66		49	59	
Hispanic (%)	a	a		a	a		a	a		a	a	
Race--White (%)	78	83		74	67		55	59		83	86	
Race--African American (%)	22	17		26	33		45	41		18	14	
Race--Other (%)	0	0		0	0		0	0		0	0	
Family income less than $30,000 (%)	41	50		51	54		73	47	*	34	52	
Family income between $30,000 and $60,000 (%)	43	33		39	34		23	36		47	36	
Family income over $60,000 (%)	16	17		11	12		4	18	*	19	12	
Eligible for free or reduced price lunch (%)	42	46		52	38	*	56	42	*	46	46	
Has any learning or other disability (%)	27	36		27	30		32	30		28	29	
Mother has bachelor's degree or higher (%)	12	17		5	9		0	17		13	24	*
Screening Tests												
TOWRE Sight Word Efficiency	84.1	85.3		84.1	85.3		83.5	84.8		82.9	83.4	
TOWRE Phonemic Decoding Efficiency	81.7	79.8	*	78.6	79.7		82.7	82.2		80.0	80.5	
Peabody Picture Vocabulary Test--Revised	94.8	95.1		92.0	92.4		91.6	100.0	*	95.2	97.3	
Baseline Tests												
WRM Word Identification	90.4	89.0		87.4	87.6		87.9	90.0		87.6	89.3	
TOWRE Phonemic Decoding Efficiency	82.1	81.8		78.0	79.9		82.9	81.2		80.6	81.3	
WRM Word Attack	93.5	92.7		91.0	92.2		93.3	94.4		93.6	93.7	
TOWRE Sight Word Efficiency	84.2	85.7		83.3	85.5		84.2	84.6		83.4	82.9	
AIMSWeb (Raw score)	79.0	75.6		80.2	79.3		74.7	80.2	*	75.0	71.6	
WRM Passage Comprehension	92.4	92.2		91.6	92.8		90.5	96.4	*	92.0	93.3	
GRADE	91.4	92.2		89.9	89.1		92.0	95.2	*	88.1	91.5	
Woodcock Johnson Spelling	93.8	92.2		90.1	91.3		91.1	92.3		88.3	86.9	
Woodcock Johnson Calculation	94.0	93.5		94.9	94.5		93.6	95.2		94.5	94.5	
Sample Size	62	66		56	46		54	38		55	31	

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

Note: All test scores are shown as standard scores, unless otherwise indicated. 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.

* Difference between treatment and control groups is statistically significant at the 0.05 level.

Difference across the four interventions (with treatment and control groups pooled within each intervention) is statistically significant at the 0.05 level.

a Values suppressed to protect student confidentiality.

Baseline Means	Treatment	Control	Treatment	Control	
Student Characteristics					
Age	8.7	8.6	8.7	8.6	
Male (%)	56	47	57	42	*

Table G.4

Baseline Characteristics of Full Sample and Three Word-level Interventions, by Treatment Status,
Analysis Sample: 5th Grade (PSSA Sample)

Baseline Means	All Interventions Treatment	All Interventions Control		Interventions Treatment	Interventions Control	
Student Characteristics						
Age	10.7	10.6	*	10.8	10.6	*
Male (%)	53	58		53	61	
Hispanic (%)	a	a		a	a	
Race--White (%)	73	74		71	71	
Race--African American (%)	27	26		29	29	
Race--Other (%)	a	a		a	a	
Family income less than $30,000 (%)	49	51		51	51	
Family income between $30,000 and $60,000 (%)	39	35		37	35	
Family income over $60,000 (%)	13	14		12	14	
Eligible for free or reduced price lunch (%)	49	43		51	42	*
Has any learning or other disability (%)	28	31		29	30	
Mother has bachelor's degree or higher (%)	8	17	*	6	17	*
Screening Tests						
TOWRE Sight Word Efficiency	83.7	84.7		83.5	84.5	
TOWRE Phonemic Decoding Efficiency	80.7	80.5		80.4	80.8	
Peabody Picture Vocabulary Test--Revised	93.4	96.1	*	93.0	96.5	*
Baseline Tests						
WRM Word Identification	88.3	89.0		87.6	89.0	
TOWRE Phonemic Decoding Efficiency	80.8	81.1		80.4	80.8	
WRM Word Attack	92.9	93.2		92.6	93.4	
TOWRE Sight Word Efficiency	83.8	84.7		83.6	84.3	
AIMSWeb (Raw score)	77.3	76.6		76.7	76.9	
WRM Passage Comprehension	91.7	93.6		91.4	94.1	*
GRADE	90.3	91.9		89.9	91.8	*
Woodcock Johnson Spelling	90.8	90.6		89.8	90.1	
Woodcock Johnson Calculation	94.3	94.4		94.3	94.7	
Sample Size	227	181		165	115	

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

Note: All test scores are shown as standard scores, unless otherwise indicated. 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.

* Difference between treatment and control groups is statistically significant at the 0.05 level.

a Values suppressed to protect student confidentiality.

APPENDIX H

Table H.1

Standard Errors of Impact Estimates for 3rd and 5th Grade Cohorts

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	1.04	1.15	1.90	1.91	1.56	2.33
TOWRE PDE	0.98	1.15	1.78	0.85	1.33	2.82
Word Identification	0.59	0.67	0.98	1.00	1.00	1.36
TOWRE SWE	0.54	0.68	0.87	0.90	1.01	1.67
AIMSweb	2.79	3.54	2.84	2.63	6.74	6.66
Passage Comprehension	0.85	1.06	1.07	1.41	2.40	1.47
GRADE	1.78	2.11	3.45	3.00	5.20	2.78

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	0.67	0.88	0.91	1.96	1.78	0.98
TOWRE PDE	0.90	1.05	1.42	2.00	1.65	1.62
Word Identification	0.68	0.87	0.65	1.23	1.68	1.54
TOWRE SWE	0.86	0.95	1.39	1.15	1.11	1.92
AIMSweb	1.84	1.97	3.86	2.13	3.96	3.53
Passage Comprehension	0.78	0.91	1.34	1.19	1.39	1.78
GRADE	0.96	1.26	1.22	2.23	2.71	1.43

Table H.2

Standard Errors of Impact Estimates for 3rd and 5th Grade Cohorts with Low Baseline Word Attack Scores

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	1.71	2.02	2.98	3.34	3.47	3.72
TOWRE PDE	1.76	2.02	3.25	3.33	3.39	3.73
Word Identification	0.96	1.13	1.61	1.79	1.93	2.05
TOWRE SWE	1.02	1.20	1.76	1.92	2.13	2.08
AIMSweb	3.41	4.02	5.69	6.28	6.69	7.63
Passage Comprehension	1.73	2.02	3.10	3.30	3.25	3.90
GRADE	2.51	2.95	4.19	4.63	5.10	5.44

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	1.37	1.56	2.62	2.39	2.60	2.73
TOWRE PDE	1.36	1.54	2.61	2.36	2.67	2.73
Word Identification	1.07	1.22	2.01	1.86	2.09	2.20
TOWRE SWE	1.00	1.14	1.87	1.70	1.90	2.06
AIMSweb	2.91	3.32	5.44	4.99	5.43	6.05
Passage Comprehension	1.49	1.68	2.84	2.53	2.91	2.96
GRADE	2.06	2.35	3.94	3.56	4.06	4.11

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	2.19	2.55	4.06	3.84	4.45	4.42
TOWRE PDE	0.91	1.07	1.80	1.58	1.81	1.76
Word Identification	1.14	1.33	2.17	1.96	2.30	2.25
TOWRE SWE	0.93	1.07	1.81	1.62	1.83	1.74
AIMSweb	3.11	3.61	6.03	5.31	6.34	6.03
Passage Comprehension	1.20	1.37	2.29	2.13	2.27	2.28
GRADE	2.38	2.76	4.68	3.93	4.82	4.59

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	1.14	1.30	2.22	2.29	2.31	1.98
TOWRE PDE	1.08	1.22	2.15	2.20	2.19	1.88
Word Identification	1.11	1.25	2.20	2.20	2.23	1.87
TOWRE SWE	0.95	1.08	1.91	1.93	1.97	1.67
AIMSweb						

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	1.86	2.09	3.41	3.43	3.49	3.38
TOWRE PDE	1.53	1.64	3.01	2.68	2.72	2.45
Word Identification	1.09	1.19	1.94	1.86	1.83	1.88
TOWRE SWE	1.19	1.30	2.11	2.02	1.96	2.09
AIMSweb	4.68	5.22	8.54	8.63	8.95	8.55
Passage Comprehension	1.46	1.60	2.53	2.40	2.49	2.49
GRADE	2.97	3.30	5.24	5.27	5.37	5.32

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	1.65	1.90	2.42	2.27	2.89	3.04
TOWRE PDE	1.64	1.90	2.53	2.35	3.13	3.11
Word Identification	1.27	1.46	1.99	1.80	2.34	2.34
TOWRE SWE	1.27	1.46	1.93	(4-5(r)-12C)42(1)-4(.96)-494(22.)10(42)-4961(2.2)5(7)-4961(2		

Table H.7

Standard Errors of Impact Estimates for 3rd and 5th Grade Cohorts With Low Baseline Word Attack and High Peabody Picture Vocabulary Test Scores

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	2.34	2.76	3.94	4.28	4.59	5.10
TOWRE PDE	1.81	2.15	3.04	3.28	3.67	4.13
Word Identification	1.42	1.68	2.38	2.67	2.87	3.00
TOWRE SWE	1.99	2.33	3.51	3.72	4.00	4.34
AIMSweb	5.02	5.88	8.97	9.69	9.86	10.51
Passage Comprehension	2.56	3.02	4.29	4.61	4.96	5.91
GRADE	3.47	4.07	5.93	6.76	6.69	7.19

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	1.75	2.00	3.22	3.23	3.40	3.50
TOWRE PDE	1.59	1.81	2.91	2.78	3.22	3.05
Word Identification	1.44	1.67	2.44	2.48	2.96	2.90
TOWRE SWE	1.42	1.62	2.56	2.40	2.91	2.75
AIMSweb	4.44	5.10	8.20	8.33	8.96	8.67
Passage Comprehension	2.35	2.67	4.34	4.24	4.85	4.31
GRADE	3.39	3.87	6.36	6.19	7.20	6.35

Table H.8

Standard Errors of Impact Estimates for 3rd and 5th Grade Cohorts with High Baseline Word Attack and High Screening Peabody Picture Vocabulary Test Scores

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	1.74	2.07	2.98	2.53	4.34	3.02
TOWRE PDE	1.41	1.66	2.42	2.17	3.24	2.56
Word Identification	1.51	1.81	2.55	2.34	3.79	2.72
TOWRE SWE	1.45	1.73	2.47	2.26	3.51	2.60
AIMSweb	4.35	5.11	7.42	6.72	10.08	8.12
Passage Comprehension	1.64	1.91	2.87	2.61	3.70	2.94
GRADE	3.60	4.29	6.13	5.43	8.43	6.75

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	1.54	1.87	2.80	3.39	3.64	2.46
TOWRE PDE	1.50	1.81	2.80	3.40	3.51	2.41
Word Identification	1.32	1.62	2.37	3.14	2.98	2.04
TOWRE SWE	1.33	1.61	2.49	3.00	3.10	2.11
AIMSweb	3.69	4.46	6.94	8.06	8.74	6.01
Passage Comprehension	1.80	2.16	3.42	3.92	4.30	2.85
GRADE	2.08	2.58	3.64	4.88	5.08	3.25

Table H.9

Standard Errors of Impact Estimates for 3rd and 5th Grade Cohorts Eligible for Free or Reduced Price School Lunch

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	1.89	2.04	2.84	2.92	2.77	3.21
TOWRE PDE	1.31	1.39	2.01	1.89	1.84	2.12
Word Identification	1.53	1.70	2.46	2.65	2.39	2.91
TOWRE SWE	1.43	1.54	2.25	2.23	2.15	2.50
AIMSweb	4.52	4.96	7.01	7.50	7.00	8.30
Passage Comprehension	1.55	1.66	2.43	2.37	2.27	2.66
GRADE	3.75	4.17	5.97	6.39	5.94	7.20

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
Word Attack	1.79	2.04	2.74	2.96	3.15	3.09
TOWRE PDE	1.58	1.80	2.43	2.62	2.78	2.78
Word Identification	1.25	1.41	1.85	2.02	2.18	2.08
TOWRE SWE	1.16	1.31	1.65	1.83	1.90	1.88
AIMSweb	3.57	4.01	5.48	5.68	6.08	6.02
Passage Comprehension	1.61	1.80	2.42	2.50	2.76	2.71
GRADE	2.00	2.24	2.86	3.09	3.25	3.32

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	26.53	0.00	22.77	0.00	6.77	0.01	8.11	0.00	13.59	0.00	5.04	0.02
TOWRE PDE	16.38	0.00	17.76	0.00	0.51	0.50	33.24	0.00	9.71	0.00	3.82	0.05
Word Identification	15.28	0.00	13.96	0.00	3.44	0.06	0.48	0.50	17.03	0.00	3.71	0.05
TOWRE SWE	10.05	0.00	5.64	0.02	5.25	0.02	0.90	0.50	6.50	0.01	0.67	0.50
AIMSweb	3.65	0.05	1.60	0.20	7.75	0.01	5.15	0.02	0.28	0.50	0.34	0.50
Passage Comprehension	6.01	0.01	1.53	0.21	17.03	0.00	0.01	0.50	2.13	0.14	0.05	0.50
GRADE	0.30	0.50	0.03	0.50	0.66	0.50	0.48	0.50	0.00	0.50	0.15	0.50

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	15.90	0.00	18.89	0.00	0.75	0.50	3.17	0.07	19.32	0.00	0.03	0.50
TOWRE PDE	3.66DEE											

Table H.13

Test Statistics and P-values of Impacts for 3rd and 5th Grade Cohorts with High Baseline Word Attack Scores

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	13.86	0.00	10.77	0.00	3.99	0.04	1.39	0.24	4.23	0.04	5.15	0.02
TOWRE PDE	36.89	0.00	54.85	0.00	0.00	0.50	39.69	0.00	51.66	0.00	5.34	0.02
Word Identification	15.22	0.00	9.90	0.00	5.27	0.02	0.35	0.50	6.66	0.01	3.64	0.05
TOWRE SWE	32.37	0.00	17.98	0.00	22.90	0.00	1.05	0.31	6.49	0.01	9.74	0.00
AIMSweb	4.23	0.04	4.82	0.03	0.02	0.50	2.35	0.12	5.02	0.02	0.49	0.50
Passage Comprehension	5.17	0.02	0.56	0.50	11.10	0.00	0.94	0.50	0.26	0.50	0.80	0.50
GRADE	0.08	0.50	1.40	0.23	4.74	0.03	3.21	0.07	14.05	0.00	0.16	0.50

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	1.03	0.31	1.50	0.22	0.01	0.50	0.49	0.50	22.89	0.00	0.48	0.50
TOWRE PDE	2.49	0.11	1.49	0.22	0.89	0.50	0.95	0.50	0.39	0.50	0.04	0.50
Word Identification	4.75	0.03	2.81	0.09	3.01	0.08	1.68	0.19	0.18	0.50	2.56	0.11
TOWRE SWE	1.82	0.17	1.80	0.18	0.24	0.50	15.44	0.00	0.18	0.50	0.04	0.50
AIMSweb	0.46	0.50	0.26	0.50	0.24	0.50	1.32	0.25	0.01	0.50	2.64	0.10
Passage Comprehension	5.60	0.02	3.41	0.06	2.29	0.13	1.91	0.16	0.00	0.50	4.11	0.04
GRADE	2.97	0.08	0.71	0.50	6.00	0.01	0.01	0.50	0.14	0.50	4.20	0.04

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table H.15

	All Interventions ABCD		Word-level Interventions BCD		Spell Read A				Corrective Reading C			
							B				D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	11.92	0.00	11.32	0.00	1.59	0.21	18.65	0.00	2.33	0.12	5.05	0.02
TOWRE PDE	9.81	0.00	8.24	0.00	1.34	0.25	47.17	0.00	2.25	0.13	0.00	0.50
Word Identification	12.42	0.00	8.88	0.00	5.35	0.02	1.31	0.25	15.58	0.00	0.88	0.50
TOWRE SWE	4.26	0.04	1.90	0.17	6.12	0.01	1.17	0.28	1.42	0.23	0.28	0.50
AIMSweb	14.25	0.00	10.87	0.00	7.22	0.01	4.95	0.02	16.65	0.00	0.44	0.50
Passage Comprehension	5.51	0.02	1.77	0.18	17.11	0.00	0.07	0.50	3.25	0.07	0.17	0.50
GRADE	1.86	0.17	0.56	0.50	2.58	0.10	0.90	0.50	0.15	0.50	1.05	0.31
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	4.82	0.03	6.07	0.01	0.19	0.50	7.91	0.01	3.08	0.08	0.52	0.50
TOWRE PDE	5.81	0.02	12.45	0.00	0.40	0.50	21.60	0.00	2.24	0.13	0.71	0.50
Word Identification	0.78	0.50	0.54	0.50	0.46	0.50	1.97	0.16	1.41	0.23	1.73	0.19
TOWRE SWE	1.10	0.29	0.57	0.50	0.85	0.50	4.35	0.03	0.80	0.50	1.67	0.19
AIMSweb	9.36	0.00	8.13	0.00	1.48	0.22	0.88	0.50	1.92	0.16	11.20	0.00
Passage Comprehension	3.40	0.06	1.83	0.17	1.84	0.17	0.59	0.50	0.71	0.50	0.55	0.50
GRADE	0.03	0.50	0.06	0.50	1.71	0.19	0.08	0.50	1.18	0.28	0.78	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	11.31	0.00	11.75	0.00	0.21	0.50	2.36	0.12	3.30	0.07	4.66	0.03
TOWRE PDE	0.98	0.50	0.59	0.50	0.36	0.50	1.29	0.25	2.52	0.11	1.51	0.22
Word Identification	0.94	0.50	0.32	0.50	0.62	0.50	0.77	0.50	0.73	0.50	2.19	0.14
TOWRE SWE	0.30	0.50	0.22	0.50	0.11	0.50	3.53	0.06	1.48	0.22	0.08	0.50
AIMSweb	5.53	0.02	7.81	0.01	1.43	0.23	0.48	0.50	13.06	0.00	0.74	0.50
Passage Comprehension	0.30	0.50	0.00	0.50	1.52	0.22	1.06	0.30	0.30	0.50	0.03	0.50
GRADE	0.13	0.50	0.39	0.50	3.40	0.06	0.77	0.50	1.08	0.30	0.70	0.50

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	8.91	0.00	12.23	0.00	0.30	2.10	0.25	0.22	4-2(.55	9n8	01(	317(0	0(50 | 2.t11 | 53 | -7.8. | 06 | 50).5	48	0gy7	22	699.5	08	0.50

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	0.00	0.50	0.01	0.50	0.01	0.50	2.02	0.15	0.61	0.50	6.15	0.01
TOWRE PDE	2.45	0.11	1.70	0.19	0.49	0.50	14.62	0.00	1.33	0.25	0.65	0.50
Word Identification	5.50	0.02	3.50	0.06	3.35	0.06	1.31	0.25	25.55	0.00	0.07	0.50
TOWRE SWE	0.96	0.50	0.48	0.50	2.21	0.13	0.01	0.50	3.38	0.06	0.19	0.50
AIMSweb	5.76	0.02	4.13	0.04	4.10	0.04	2.73	0.09	3.62	0.05	0.14	0.50
Passage Comprehension	3.83	0.05	2.23	0.13	8.51	0.00	0.20	0.50	3.24	0.07	1.34	0.25
GRADE	1.46	0.22	0.12	0.50	5.34	0.02	0.83	0.50	0.79	0.50	0.10	0.50

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	14.84	0.00	17.44	0.00	0.55	0.50	6.98	0.01	11.68	0.00	0.39	0.50
TOWRE PDE	12.20	0.00	25.00	0.00	0.80	0.50	18.29	0.00	4.			

Table H.18

Test Statistics and P-values of Impacts for 3rd and 5th Grade Cohorts with High Baseline Word Attack and High Screening Peabody Picture Vocabulary Test Scores

	All Interventions ABCD		Word-level Interventions BCD		Failure Free Reading A		Spell Read B		Wilson Reading C		Corrective Reading D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	59.88	0.00	49.59	0.00	3.81	0.05	64.26	0.00	3.86	0.05	10.49	0.00
TOWRE PDE	10.36	0.00	14.56	0.00	0.95	0.50	19.70	0.00	5.61	0.02	1.37	0.24
Word Identification	4.01	0.04	6.21	0.01	0.41	0.50	1.69	0.19	0.24	0.50	7.11	0.01
TOWRE SWE	19.19	0.00	12.51	0.00	8.16	0.00	0.01	0.50	6.71	0.01	8.93	0.00
AIMSweb	3.02	0.08	3.63	0.05	0.00	0.50	1.02	0.31	10.23	0.00	0.00	0.50
Passage Comprehension	7.11	0.01	0.70	0.50	17.88	0.00	0.03	0.50	0.20	0.50	0.70	0.50
GRADE	0.25	0.50	0.21	0.50	0.07	0.50	2.18	0.14	1.12	0.29	0.14	0.50

	All Interventions ABCD		Word-level Interventions BCD		Failure Free Reading A		Spell Read B		Wilson Reading C		Corrective Reading D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	1.02	0.31	1.20	0.27	0.00	0.50	1.04	0.31	1.43	0.23	0.74	0.50
TOWRE PDE	4.55	0.03	4.42	0.03	0.20	0.50	1.62	0.20	3.03	0.08	0.78	0.50
Word Identification	2.80	0.09	2.71	0.10	0.10	0.50	4.84	0.03	0.02	0.50	1.22	0.27
TOWRE SWE	2.27	0.13	1.25	0.26	1.27	0.26	0.45	0.50	2.52	0.11	0.63	0.50
AIMSweb	0.70	0.50	0.73	0.50	0.02	0.50	0.08	0.50	0.23	0.50	9.38	0.00
Passage Comprehension	0.46	0.50	0.03	0.50	2.19	0.13	0.01	0.50	0.03	0.50	0.25	0.50
GRADE	6.68	0.01	4.11	0.04	6.46	0.01	1.56	0.21	0.35	0.50	7.13	0.01

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table H.19

Test Statistics and P-values of Impacts for 3rd and 5th Grade Cohorts Eligible for Free or Reduced Price School Lunch

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	9.30	0.00	4.17	0.04	7.65	0.01	0.49	0.50	10.85	0.00	0.83	0.50
TOWRE PDE	6.07	0.01	7.08	0.01	1.22	0.27	16.30	0.00	7.38	0.01	0.16	0.50
Word Identification	3.60	0.05	3.51	0.06	1.34	0.24	0.02	0.50	7.48	0.01	2.77	0.09
TOWRE SWE	5.05	0.02	2.36	0.12	5.54	0.02	0.87	0.50	4.61	0.03	0.00	0.50
AIMSweb	0.75	0.50	0.21	0.50	1.40	0.23	1.21	0.27	0.00	0.50	0.05	0.50
Passage Comprehension	0.02	0.50	0.01	0.50	0.01	0.50	0.06	0.50	0.03	0.50	0.00	0.50
GRADE	0.16	0.50	0.04	0.50	1.30	0.25	0.00	0.50	0.07	0.50	0.70	0.50

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	7.41	0.01	8.30	0.00	0.00	0.50	1.27	0.26	5.36	0.02	2.16	0.14
TOWRE PDE	1.80	0.18	2.07	0.15	0.00	0.50	0.06	0.50	0.02	0.50	5.96	0.01
Word Identification	0.03	0.50	0.01	0.50	0.78	0.50	0.43	0.50	0.08	0.50	0.11	0.50
TOWRE SWE	9.44	0.00	12.54	0.00	0.76	0.50	15.91	0.00	1.82	0.17	10.86	0.00
AIMSweb	0.51	0.50	3.18	0.07	2.76	0.09	1.12	0.29	1.07	0.30	1.30	0.25
Passage Comprehension	1.30	0.25	0.15	0.50	3.38	0.06	1.82	0.17	0.40	0.50	0.04	0.50
GRADE	9.04	0.00	4.88	0.03	12.91	0.00	24.78	0.00	0.03	0.50	0.03	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table H.20

Test Statistics and P-values of Impacts for 3rd and 5th Grade Cohorts Not Eligible for Free or Reduced Price School Lunch

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	6.65	0.01	6.30	0.01	0.33	0.50	7.43	0.01	1.29	0.25	5.99	0.01
TOWRE PDE	68.55	0.00	88.03	0.00	0.07	0.50	18.23	0.00	77.03	0.00	24.12	0.00
Word Identification	7.29	0.01	4.21	0.04	6.06	0.01	0.62	0.50	2.15	0.14	1.56	0.21
TOWRE SWE	3.93	0.04	2.57	0.10	4.57	0.03	0.01	0.50	1.17	0.28	2.66	0.10
AIMSweb	7.76	0.01	5.63	0.02	9.79	0.00	0.15	0.50	5.10	0.02	0.46	0.50
Passage Comprehension	5.91	0.01	2.81	0.09	21.48	0.00	0.07	0.50	2.82	0.09	0.50	0.50
GRADE	0.02	0.50	0.00	0.50	0.05	0.50	0.55	0.50	0.57	0.50	0.37	0.50

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	10.95	0.00	14.84	0.00	0.42	0.50	9.52	0.00	18.02	0.00	0.50	0.50
TOWRE PDE	5.11	0.02	8.28	0.00	0.00	0.50	9.93	0.00	4.09	0.04	0.04	0.50
Word Identification	1.01	0.32	1.15	0.28	0.00	0.50	0.01	0.50	0.03	0.50	2.79	0.09
TOWRE SWE	2.67	0.10	1.49	0.22	2.37	0.12	2.87	0.09	3.85	0.05	0.68	0.50
AIMSweb	3.06	0.08	4.32	0.04	0.00	0.50	5.80	0.02	0.14	0.50	8.18	0.00
Passage Comprehension	1.46	0.22	0.83	0.50	0.76	0.50	0.08	0.50	0.38	0.50	0.44	0.50
GRADE	0.30	0.50	0.39	0.50	0.00	0.50	0.36	0.50	0.45	0.50	1.10	0.29

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table H.22

Tests for Differences in Impacts for 3rd and 5th Grade Cohorts with Low Baseline Word Attack Scores

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	1.03	0.31	0.14	0.50	1.74	0.18	0.11	0.50	0.16	0.50	0.51	0.50
TOWRE PDE	0.03	0.50	0.00	0.50	0.23	0.50	0.00	0.50	7.42	0.01	1.39	0.24
Word Identification	1.82	0.17	1.24	0.27	0.29	0.50	0.09	0.50	0.52	0.50	0.52	0.50
TOWRE SWE	5.51	0.02	4.62	0.03	0.84	0.50	0.00	0.50	0.05	0.50	28.44	0.00
AIMSweb	0.52	0.50	2.85	0.09	1.73	0.19	0.70	0.50	2.60	0.10	2.57	0.10
Passage Comprehension	0.06	0.50	0.12	0.50	1.67	0.19	0.86	0.50	1.30	0.25	0.10	0.50
GRADE	0.50	0.50	2.55	0.11	1.79	0.18	7.53	0.01	19.58	0.00	0.05	0.50

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	5.58	0.02	8.25	0.00	0.06	0.50	7.67	0.01	0.23	0.50	3.06	0.08
TOWRE PDE	0.42	0.50	2.21	0.13	2.60	0.10	0.32	0.50	0.49	0.50	2.98	0.08
Word Identification	1.56	0.21	0.71	0.50	3.23	0.07	0.95	0.50	0.48	0.50	1.57	0.21
TOWRE SWE	1.10	0.29	0.57	0.50	1.04	0.31	0.12	0.50	0.11	0.50	0.47	0.50
AIMSweb	1.47	0.22	1.22	0.27	0.45	0.50	2.96	0.08	2.10	0.14	0.75	0.50
Passage Comprehension	3.06	0.08	2.85	0.09	0.36	0.50	1.08	0.30	0.91	0.50	1.09	0.30
GRADE	6.48	0.01	5.89	0.01	1.13	0.29	0.14	0.50	7.91	0.01	4.51	0.03

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	0.03	0.50	0.68	0.50	2.00	0.15	4.35	0.03	0.01	0.50	0.68	0.50
TOWRE PDE	0.25	0.50	0.23	0.50	0.11	0.50	1.89	0.17	0.14	0.50	6.05	0.01
Word Identification	0.33	0.50	0.50	0.50	0.00	0.50	1.67	0.19	0.02	0.50	0.00	0.50
TOWRE SWE	0.65	0.50	0.22	0.50	0.91	0.50	0.07	0.50	0.02	0.50	0.15	0.50
AIMSweb	5.72	0.02	8.10	0.00	0.22	0.50	2.09	0.14	7.43	0.01	0.49	0.50
Passage Comprehension	1.81	0.17	0.85	0.50	1.99	0.16	0.08	0.50	3.68	0.05	0.03	0.50
GRADE	0.72	0.50	0.00	0.50	4.06	0.04	0.04	0.50	2.28	0.13	2.17	0.14

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	0.03	0.50	0.21	0.50	0.39	0.50	2.68	0.10	0.40	0.50	2.79	0.09
TOWRE PDE	0.03	0.50	0.91	0.50	0.89	0.50	5.62	0.02	0.28	0.50	1.86	0.17
Word Identification	0.12	0.50	0.03	0.50	0.35	0.50	13.35	0.00	1.79	0.18	0.01	0.50
TOWRE SWE	1.00	0.32	1.63	0.20	0.28	0.50	0.21	0.50	0.07	0.50	4.98	0.02
AIMSweb	0.04	0.50	0.08	0.50	2.41	0.12	0.38	0.50	0.84	0.50	0.01	0.50
Passage Comprehension	5.01	0.02	5.16	0.02	0.45	0.50	1.01	0.32	4.10	0.04	1.58	0.21
GRADE	2.28	0.13	0.81	0.50	4.89	0.03	0.04	0.50	0.00	0.50	3.18	0.07

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p*

Table H.24

Tests for Differences in Impacts for 3rd and 5th Grade Cohorts with Low Baseline Word Attack and Low Screening Peabody Picture Vocabulary Test Scores

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	0.61	0.50	0.00	0.50	2.93	0.08	0.00	0.50	0.71	0.50	0.98	0.50
TOWRE PDE	0.41	0.50	0.64	0.50	0.01	0.50	5.77	0.02	0.00	0.50	1.07	0.30
Word Identification	3.11	0.07	3.18	0.07	0.13	0.50	1.46	0.23	1.96	0.16	0.18	0.50
TOWRE SWE	4.15	0.04	2.48	0.11	3.37	0.06	1.35	0.24	3.60	0.05	0.68	0.50
AIMSweb	19.23	0.00	17.77	0.00	0.55	0.50	3.30	0.07	14.10	0.00	0.00	0.50
Passage Comprehension	2.43	0.12	1.37	0.24	1.29	0.25	0.71	0.50	1.33	0.25	0.00	0.50
GRADE	2.03	0.15	0.10	0.50	12.45	0.00	0.02	0.50	2.40	0.12	0.98	0.50

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	0.07	0.50	0.52	0.50	0.56	0.50	0.82	0.50	7.96	0.01	6.95	0.01
TOWRE PDE	0.73	0.50	0.23	0.50	1.06	0.30	1.78	0.18	0.11	0.50	1.84	0.17
Word Identification	3.11	0.07	3.18	0.07	0.13	0.50	1.46	0.23	1.96	0.16	0.18	0.50
TOWRE SWE	1.07	0.30	0.95	0.50	0.22	0.50	0.01	0.50	0.14	0.50	1.82	0.17
AIMSweb	1.89	0.17	2.36	0.12	0.04	0.50	1.54	0.21	0.33	0.50	1.93	0.16
Passage Comprehension	12.88	0.00	12.19	0.00	1.30	0.25	17.05	0.00	0.58	0.50	7.73	0.01
GRADE	0.54	0.50	0.01	0.50	6.95	0.01	0.65	0.50	0.01	0.50	0.55	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table H.25

Tests for Differences in Impacts for 3rd and 5th Grade Cohorts With Low Baseline Word Attack and High Peabody Picture Vocabulary Test Scores

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	4.10	0.04	3.01	0.08	1.37	0.24	0.10	0.50	0.09	0.50	10.78	0.00
TOWRE PDE	2.83	0.09	6.08	0.01	0.79	0.50	0.32	0.50	0.80	0.50	9.68	0.00
Word Identification	0.00	0.50	0.02	0.50	0.09	0.50	2.05	0.15	1.55	0.21	0.96	0.50
TOWRE SWE	0.02	0.50	0.01	0.50	0.01	0.50	0.10	0.50	1.05	0.31	0.77	0.50
AIMSweb	1.99	0.15	2.33	0.12	0.00	0.50	1.29	0.25	2.07	0.15	0.16	0.50
Passage Comprehension	0.72	0.50	1.11	0.29	0.20	0.50	0.33	0.50	2.87	0.09	0.20	0.50
GRADE	1.33	0.25	0.05	0.50	8.86	0.00	1.30	0.25	1.25	0.26	0.19	0.50

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	3.45	0.06	4.18	0.04	0.02	0.50	3.03	0.08	1.42	0.23	0.26	0.50
TOWRE PDE	3.44	0.06	10.01	0.00	1.87	0.17	7.17	0.01	2.25	0.13	0.62	0.50
Word Identification	0.12	0.50	0.81	0.50	4.55	0.03	8.59	0.00	0.01	0.50	0.03	0.50
TOWRE SWE	0.85	0.50	0.95	0.50	0.08	0.50	0.78	0.50	0.85	0.50	0.02	0.50
AIMSweb	0.72	0.50	0.27	0.50	1.14	0.28	0.03	0.50	4.16	0.04	0.34	0.50
Passage Comprehension	0.07	0.50	0.03	0.50	0.04	0.50	1.19	0.27	0.76	0.50	0.08	0.50
GRADE	7.54	0.01	10.67	0.00	0.01	0.50	1.48	0.22	5.20	0.02	3.44	0.06

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	1.88	0.17	2.98	0.08	0.19	0.50	12.13	0.00	0.04	0.50	0.09	0.50
TOWRE PDE	0.31	0.50	0.14	0.50	0.34	0.50	0.09	0.50	0.00	0.50	0.48	0.50
Word Identification	0.31	0.50	0.00	0.50	1.76	0.18	0.18	0.50	1.72	0.19	0.61	0.50
TOWRE SWE	3.43	0.06	3.24	0.07	0.20	0.50	0.20	0.50	0.23	0.50	19.42	0.00
AIMSweb	0.02	0.50	1.07	0.30	3.10	0.07	0.08	0.50	3.33	0.06	0.08	0.50
Passage Comprehension	0.46	0.50	0.02	0.50	3.00	0.08	0.54	0.50	0.46	0.50	0.05	0.50
GRADE	0.10	0.50	0.01	0.50	0.65	0.50	1.97	0.16	2.68	0.10	0.29	0.50

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack												

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
Word Attack	0.79	0.50	2.23	0.13	2.75	0.09	1.97	0.16	0.44	0.50	0.93	0.50
TOWRE PDE	13.02	0.00	25.75	0.00	0.42	0.50	0.01	0.50	35.69	0.00	16.66	0.00
Word Identification	0.20	0.50	0.36	0.50	0.02	0.50	0.14	0.50	0.30	0.50	1.03	0.31
TOWRE SWE												

Table H.29

Standard Errors of Impact Estimates for 3rd and 5th Grade Cohorts with Low Baseline Word Attack Scores

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	27.93	33.30	44.69	46.15	56.25	70.78
PSSA Math	32.04	38.18	49.73	51.14	69.13	77.86

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	21.28	24.15	40.32	35.35	41.50	40.20
PSSA Math	23.10	26.26	43.54	40.23	45.20	43.84

Table H.30

Standard Errors of Impact Estimates for 3rd and 5th Grade Cohorts with High Baseline Word Attack Scores

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	22.87	26.31	44.17	37.61	47.74	42.27
PSSA Math	21.80	25.01	44.47	38.03	43.20	40.73

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	27.74	31.19	55.86	51.66	59.66	49.10
PSSA Math	24.25	27.63	47.69	47.30	51.19	43.47

Table H.31

Standard Errors of Impact Estimates for 3rd and 5th Grade Cohorts with Low Screening Peabody Picture Vocabulary Test

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	34.14	36.41	55.12	49.44	55.10	57.75
PSSA Math	47.59	52.91	82.63	80.88	86.50	89.34

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	29.66	34.13	43.38	38.28	52.33	55.03
PSSA Math	30.01	34.45	44.12	38.46	52.67	54.88

All Interventions	Word-level Interventions	Failure Free

Table H.33

Standard Errors of Impact Estimates for 3rd and 5th Grade Cohorts with Low Baseline Word Attack and Low Screening Peabody Picture Vocabulary Test Scores

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	40.02	44.24	64.46	61.58	87.25	78.74
PSSA Math	60.15	68.31	100.33	99.42	136.83	123.52

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	42.17	49.91	64.21	58.25	73.79	77.05
PSSA Math	43.09	51.42	64.55	59.10	75.71	78.11

Table H.34

Standard Errors of Impact Estimates for 3rd and 5th Grade Cohorts With Low Baseline Word Attack and High Peabody Picture Vocabulary Test Scores

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	31.56	37.36	51.84	56.16	66.80	70.60
PSSA Math	30.90	36.96	49.26	56.26	67.54	67.08

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	29.54	33.81	56.98	58.27	60.90	57.45
PSSA Math	28.99	33.08	55.19	57.85	59.03	55.55

All

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	27.42	30.30	43.69	44.65	42.23	48.40
PSSA Math	30.48	33.72	45.97	49.04	46.77	53.88

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	29.80	32.95	39.73	44.25	46.23	43.35
PSSA Math						

Table H.37

Standard Errors of Impact Estimates for 3rd and 5th Grade Cohorts Not Eligible for Free or Reduced Price School Lunch

Grade 3 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	30.23	37.56	43.91	45.50	88.33	53.78
PSSA Math	35.92	45.21	47.41	48.14	111.83	55.52

Grade 5 Cohort	All Interventions Standard Error	Word-level Interventions Standard Error	Failure Free Reading Standard Error	Spell Read Standard Error	Wilson Reading Standard Error	Corrective Reading Standard Error
PSSA Reading	23.24	27.36	45.47	43.52	54.74	44.68
PSSA Math	23.25	27.07	46.94	43.62	51.40	45.97

Table H.38

Test Statistics and P-values of Impacts for 3rd and 5th Grade Cohorts

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.85	0.50	0.03	0.50	4.94	0.02	3.51	0.06	1.23	0.27	0.71	0.50
PSSA Math	1.24	0.26	0.79	0.50	0.78	0.50	0.27	0.50	6.91	0.01	0.01	0.50

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	5.39	0.02	2.99	0.08	6.09	0.01	3.15	0.07	0.67	0.50	0.54	0.50
PSSA Math	11.34	0.00	10.91	0.00	1.51	0.22	1.16	0.28	9.75	0.00	4.17	0.04

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Grade 3 Cohort	Test	*p*	Test	*p*

	All Interventions											
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.50	0.50	0.94	0.50	0.10	0.50	0.14	0.50	1.43	0.23	0.45	0.50
PSSA Math	3.43	0.06	2.82	0.09	1.28	0.26	2.06	0.15	3.88	0.05	0.21	0.50

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.02	0.50	0.01	0.50	0.00	0.50	0.97	0.50	5.98	0.01	4.81	0.03
PSSA Math	3.74	0.05	3.62	0.05	0.58	0.50	3.54	0.06	2.56	0.11	0.75	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table H.42

Test Statistics and P-values of Impacts for 3rd and 5th Grade Cohorts with High Screening Peabody Picture Vocabulary Test Scores

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.00	0.50	0.09	0.50	0.51	0.50	1.00	0.50	0.81	0.50	0.02	0.50
PSSA Math	0.22	0.50	0.58	0.50	0.39	0.50	3.11	0.07	2.54	0.11	0.15	0.50

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	3.74	0.05	1.60	0.20	4.96	0.02	0.02	0.50	0.03	0.50	5.57	0.02
PSSA Math	1.85	0.17	1.98	0.16	0.04	0.50	0.61	0.50	7.29	0.01	0.48	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	1.79	0.18	3.49	0.06	0.19	0.50	0.34	0.50	3.44	0.06	6.30	0.01
PSSA Math	5.68	0.02	5.72	0.02	1.47	0.22	1.93	0.16	4.68	0.03	1.44	0.23

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.00	0.50	0.06	0.50	0.57	0.50	0.25	0.50	5.03	0.02	11.34	0.00
PSSA Math	0.65	0.50	0.55	0.50	0.09	0.50	0.62	0.50	2.25	0.13	0.08	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) wi(a)- Tc68

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.06	0.50	0.11	0.50	3.46	0.06	0.02	0.50	0.13	0.50	0.01	0.50
PSSA Math	1.24	0.27	2.50	0.11	0.78	0.50	0.70	0.50	1.83	0.17	7.98	0.00

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	2.40	0.12	1.02	0.31	1.84	0.17	0.14	0.50	0.12	0.50	1.18	0.28
PSSA Math	0.04	0.50	0.94	0.50	0.46	0.50	0.22	0.50	6.53	0.01	0.10	0.50

Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.02	0.50	0.00	0.50	0.13	0.50	9.23	0.00	7.43	0.01	0.02	0.50
PSSA Math	0.44	0.50	0.69	0.50	0.09	0.50	6.40	0.01	1.79	0.18	6.68	0.01

Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	5.03	0.02	3.44	0.06	1.20	0.27	0.63	0.50	0.21	0.50	4.80	0.03
PSSA Math	2.06	0.15	1.66	0.19	0.47	0.50	0.00	0.50	0.96	0.50	3.62	0.05

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p*

Table H.48

Tests for Differences in Impacts for 3rd and 5th Grade Cohorts

	All Interventions ABCD		Word-level Interventions BCD		Failure Free Reading A		Spell Read B		Wilson Reading C		Corrective Reading D	
	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.33	0.50	0.72	0.40	0.44	0.50	0.13	0.50	1.89	0.17	0.00	0.50
PSSA Math	5.97	0.01	6.41	0.01	1.33	0.25	0.02	0.50	16.17	0.00	1.38	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table H.49

Tests for Differences in Impacts for 3rd and 5th Grade Cohorts with Low Baseline Word Attack Scores

	All Interventions		Word-level									
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.80	0.50	4.15	0.04	1.49	0.22	0.99	0.50	0.39	0.50	3.29	0.07
PSSA Math	6.08	0.01	7.35	0.01	0.00	0.50	0.30	0.50	2.01	0.15	15.30	0.00
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.09	0.50	0.32	0.50	0.31	0.50	0.00	0.50	0.39	0.50	2.28	0.13
PSSA Math	1.04	0.31	0.50	0.50	0.56	0.50	0.01	0.50	0.25	0.50	0.70	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table H.50

Tests for Differences in Impacts for 3rd and 5th Grade Cohorts with Low Screening Peabody Picture Vocabulary Test Scores

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.44	0.50	0.59	0.50	0.00	0.50	0.43	0.50	0.11	0.50	0.31	0.50
PSSA Math	2.44	0.11	1.56	0.21	1.51	0.22	5.34	0.02	0.29	0.50	0.14	0.50

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	1.85	0.17	0.84	0.50	1.53	0.21	0.26	0.50	0.50	0.50	13.53	0.00
PSSA Math	0.42	0.50	0.38	0.50	0.08	0.50	1.90	0.16	0.76	0.50	0.27	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table H.51

Tests for Differences in Impacts for 3rd and 5th Grade Cohorts with Low Baseline Word Attack and Low Screening Peabody Picture Vocabulary Test Scores

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	2.66	0.10	3.99	0.04	0.04	0.50	1.27	0.26	1.17	0.28	7.10	0.01
PSSA Math	5.18	0.02	4.98	0.02	1.48	0.22	2.77	0.09	1.78	0.18	1.51	0.22

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.50	0.50	0.62	0.50	0.00	0.50	0.04	0.50	1.53	0.21	19.95	0.00
PSSA Math	0.01	0.50	0.02	0.50	0.00	0.50	0.22	0.50	0.01	0.50	0.37	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table H.52

Tests for Differences in Impacts for 3rd and 5th Grade Cohorts With Low Baseline Word Attack and High Peabody Picture Vocabulary Test Scores

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.14	0.50	0.32	0.50	0.26	0.50	0.58	0.50	0.01	0.50	0.03	0.50
PSSA Math	0.03	0.50	1.41	0.23	1.78	0.18	2.14	0.14	0.23	0.50	7.56	0.01

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.03	0.50	0.10	0.50	0.17	0.50	0.21	0.50	0.13	0.50	0.08	0.50
PSSA Math	3.45	0.06	2.23	0.13	1.10	0.29	1.12	0.29	0.00	0.50	2.04	0.15

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table H.53

Tests for Differences in Impacts for 3rd and 5th Grade Cohorts with High Baseline Word Attack and High Screening Peabody Picture Vocabulary Test Scores

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 3 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	0.40	0.50	0.01	0.50	2.00	0.15	1.41	0.23	3.21	0.07	0.02	0.50
PSSA Math	3.44	0.06	2.51	0.11	0.46	0.50	0.67	0.50	0.00	0.50	8.81	0.00

	All Interventions		Word-level Interventions		Failure Free Reading		Spell Read		Wilson Reading		Corrective Reading	
	ABCD		BCD		A		B		C		D	
Grade 5 Cohort	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	1.33	0.25	1.07	0.30	0.12	0.50	0.01	0.50	0.01	0.50	4.91	0.03
PSSA Math	0.39	0.50	0.12	0.50	0.55	0.50	0.04	0.50	0.31	0.50	0.08	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

Table H.54

Tests for Differences in Impacts for 3rd and 5th Grade Cohorts Eligible for Free or Reduced Price School Lunch

Grade 3 Cohort	All Interventions ABCD		Word-level Interventions BCD		Failure Free Reading A		Spell Read B		Wilson Reading C		Corrective Reading D	
	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	7.95	0.01	3.99	0.04	8.21	0.00	1.79	0.18	1.08	0.30	5.49	0.02
PSSA Math	2.04	0.15	1.75	0.18	0.09	0.50	0.21	0.50	5.63	0.02	0.60	0.50

Grade 5 Cohort	All Interventions ABCD		Word-level Interventions BCD		Failure Free Reading A		Spell Read B		Wilson Reading C		Corrective Reading D	
	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*	Test	*p*
PSSA Reading	5.88	0.01	2.66	0.10	3.81	0.05	7.91	0.01	0.34	0.50	0.07	0.50
PSSA Math	0.43	0.50	0.03	0.50	1.94	0.16	1.18	0.28	2.69	0.10	0.16	0.50

Note: "Test" refers to the Chi-Square test statistics for a general linear hypothesis test (Wald Test) with 1 degree of freedom.
"*p* "is the *p* -value produced by HLM6 software and is automatically set to 0.5 for values that exceed 0.5.

APPENDIX I

ESTIMATED R-SQUARED VALUES AND INTRACLASS CORRELATIONS

Table I.1

Estimated R-squared

	R-squared	
Test	Grade 3 Cohort	Grade 5 Cohort
Word Attack	0.45	0.39
TOWRE PDE	0.43	0.51
Word Identification	0.48	0.55
TOWRE SWE	0.55	0.47
AIMSweb	0.57	0.69
Passage Comprehension	0.39	0.24
GRADE	0.27	0.20

Note: The R-squared is the squared correlation between baseline and second follow-up test scores.

Table I.2

Estimated Intraclass Correlations

	Intraclass Correlation	
Test	Grade 3 Cohort	Grade 5 Cohort
Word Attack	0.17	0.24
TOWRE PDE	0.17	0.12
Word Identification	0.18	0.15
TOWRE SWE	0.21	0.10
AIMSweb	0.27	0.23
Passage Comprehension	0.19	0.18
GRADE	0.25	0.23

APPENDIX J

SCIENTIFIC ADVISORY BOARD

The following individuals served on the Scientific Board of Directors of the Haan Foundation for Children, which coordinated the selection of the four chosen interventions and the funding of the interventions and evaluation.

Scientific Advisory Board
Dr. Rebecca Felton
Dr. Jackebev7eenhhedr

APPENDIX K

PSSA DATA COLLECTION FORM

SCHOOL NAME:_____

Pennsylvania System of School Assessment Scores (2003-04) for the Power4Kids Reading Initiative: Third-Grade Students

For each student listed in the Excel file sent to you, please provide the information specified in the column headings. (An example is provided in the table below.) We recommend extracting the data electronically from your files into the Excel file we provided. Alternatively, you may choose to manually enter the data into the Excel file or onto this form. Thank you.

	Name		Reading							Math		
			Points Achieved									
MPRID	Last	First	Learning to Read Independently	Reading Critically in All Content Areas	Reading, Analyzing, and Interpreting Literature	Total Reading	Scaled Score	Standard Error	Percentile	Scaled Score	Standard Error	Percentile
12345678	*Smith*	*John*	*15*	*12*	*17*	*44*	*1561*	*96*	*91*	*1539*	*79*	*88*

SCHOOL NAME:_____

Pennsylvania System of School Assessment Scores (2003-04) for the Power4Kids Reading Initiative:

APPENDIX L

SAMPLE TEST ITEMS

This appendix provides sample test items for the main tests used in the analysis.

Woodcock Reading Mastery Test-R, Word Attack Test (sample item)
Woodcock Reading Mastery Tests – Revised NU (WRMT-R/NU) Forms G and H. 1998. American Guidance Services, Inc., Circle Pines, MN.

Tester says, "I want you to read some words that are not real words. I want you to tell me how they sound."
Tester points to "tat" on the subject page.
Tester says, "How does that word sound?"

Woodcock Reading Mastery Test-R, Word Identification Test (sample item)
Woodcock Reading Mastery Tests – Revised NU (WRMT-R/NU) Forms G and H. 1998. American Guidance Services, Inc., Circle Pines, MN.

Tester points to the word "is" on the subject page.
Tester says, "What is the word?"

Woodcock Reading Mastery Test-R, Passage Comprehension Test (sample item)
Woodcock Reading Mastery Tests – Revised NU (WRMT-R/NU) Forms G and H. 1998. American Guidance Services, Inc., Circle Pines, MN.

Tester points to the sentence on the subject page and says, "This says, 'The cat is playing with a…(pause)."

Tester points to the blank space in the sentence and says, "What word belongs in the blank space?

The cat is playing with a _____.

Test of Word Reading Efficiency (TOWRE), Sight Word Efficiency Test (practice items)
Test of Word Reading Efficiency (TOWRE), Forms A and B. 1999. PRO-ED, Inc, Austin, TX.

Tester says, "I want you to read some lists of words as fast as you can. Let's start with the practice list. Begin at the top, and read down the list as fast as you can. If you come to a word you cannot read, just skip it and go to the next word."
Practice Words: on, my, bee, old, warm, bone, most, spell

“I’m waiting for Mr. Sanchez to leave for work. It shouldn’t be much longer.”

“Billy, your Mom said it was all right for you to play ball with me at the park. I don’t