FINAL REPORT

Saxon Phonics and Spelling
Randomized Control Trial

Prepared By
Miriam Resendez, M.A., Senior Researcher
Mariam Azin, Ph.D., President
October, 2007

For inquiries, please contact PRES Associates at:
info@presassociates.com
(307) 733-3255

EXECUTIVE SUMMARY

According to the National Reading Panel (2000), in order for students to read well, they need explicit, systematic instruction in five essential areas, including: 1) phonemic awareness; 2) phonics; 3) fluency; 4) vocabulary; and 5) text comprehension. At the early primary level (K-2), it is particularly important to emphasize phonics and phonemic awareness. Given the need to help students' with the skills they need to become successful readers, *Saxon* Publishers released *Saxon Phonics and Spelling*, a K-2 program designed to supplement existing classroom reading programs. In order to determine the effectiveness of the *Saxon Phonics and Spelling* program in helping students attain critical reading and spelling skills, Planning, Research, and Evaluation Services (PRES) Associates conducted a year-long study at the 1st and 2nd grades. This randomized control trial (RCT), which commenced in the Fall of 2006, was designed to fully address the quality criteria put forth by the What Works Clearinghouse (WWC).

The final sample consisted of 682 students in six geographically-dispersed schools. Teachers were randomly assigned to treatment (n=18) and control conditions (n=17).

Major findings, organized by the key evaluation questions, include:

1. ***Do phonics, reading words, and spelling skills improve over the course of participating in the Saxon Phonics and Spelling program? Does this vary across different types of students and levels of implementation?***

Students using the *Saxon Phonics and Spelling* program significantly improved over the course of the school year in the areas of spelling, phonics, and reading words, including high frequency and sight words. Gains were also observed on developmental spelling stages. In particular, among 1st grade students who took the Morris & Perney Spelling test, results showed that students moved to higher spelling stages from pre- to post-testing.

Furthermore, the *Saxon Phonics and Spelling* program worked just as well with females and males, students who spoke a language other than English at home and those that did not, and special education and non-special education students. Among the remaining subgroups, differences in improvement between students were observed. Generally, Whites, 2nd graders, students not receiving free and reduced lunch, and lower-performing students showed greater gains than minorities, 1st graders, students receiving free/reduced lunch, and higher-performing students, respectively. Nevertheless, among all subgroups, students using the *Saxon Phonics and Spelling* program showed significant gains in reading, phonics, and spelling.

Since there was some variation observed in overall implementation of the *Saxon Phonics and Spelling* program among treatment teachers, analyses were performed to examine if this affected student performance. Results showed that there was no significant relationship between overall *Saxon Phonics and Spelling* implementation levels and improved performance on the outcome measures. That is, students whose teachers used the *Saxon Phonics and Spelling* program improved on these measures, regardless of their level of overall implementation.

However, preliminary analyses of individual components of the *Saxon Phonics and*

Spelling showed that there was a relationship between teacher's use of various *Saxon Phonics and Spelling* program components (e.g., use of worksheets and fluency readers) and gains in student phonics, reading, and spelling performance. Furthermore, the percent of lessons completed in the *Saxon Phonics and Spelling* program was a strong predictor of student gains in performance – the more lessons completed, the greater the improvement.

2. ***How does phonics, reading words, and spelling performance differ between students who use Saxon Phonics and Spelling as compared to students who do not use this program? Do effects on student achievement differ across types of students or settings?***

There were notable differences in treatment and control students' performance. Students using the *Saxon Phonics and Spelling* program showed more improvement than control students on the ITBS Word Analysis (which measures phonics and phonemic awareness skills), Spelling, and Reading Words tests. Most of the effect sizes, which provide an indication of the importance of results, would be considered educationally significant by the research literature.

In addition to treatment students outperforming control students on these measures, exploratory subgroup analyses showed a number of significant differences between treatment and control students who were females, Whites, African Americans, 2nd graders, receiving free/reduced lunch, attending various schools, and lower-performing. In particular, students in these subgroups that used *Saxon Phonics and Spelling* showed greater growth in performance from pre- to post-testing as compared to students that did not use this supplemental program.

3. ***Does participation in Saxon Phonics and Spelling result in other positive student outcomes (e.g., positive attitudes towards reading and so forth)?***

While the main focus of the *Saxon Phonics and Spelling* program is to improve upon important reading and spelling skills, other measures were included to explore if *Saxon Phonics and Spelling* was associated with positive impacts on student attitudes towards reading, phonics, and so forth. Results showed that, in general, treatment and control students had similar positive attitudes in regards to phonics, spelling, writing, motivation to do well in school, and perceived reading ability. However, control students had significantly more positive attitudes towards reading as compared to treatment students during the Spring.

In terms of the program's effects on teachers, results showed notable increases in treatment teachers' levels of preparation and knowledge to teach the five elements of reading, spelling and writing, and their engagement in effective literacy practices from Fall to Spring. In addition, while during the Fall control teachers indicated having more knowledge, preparation, and engagement in effective literacy practices, treatment teachers caught up to control teachers in the areas of preparation and engagement in best practices by the Spring. Thus, there is evidence that suggests that the *Saxon Phonics and Spelling* program has a positive impact on teacher's level of preparation to teach phonics and spelling, and this in turn can lead to improvement in their pedagogical practices.

4. *What did users of the Saxon Phonics and Spelling program think about it?*

The *Saxon Phonics and Spelling* program was also highly regarded by the vast majority of teachers. A full 94% of treatment teachers surveyed agreed that that the program contributed to improved reading ability, and helped their students obtain greater phonics and spelling skills. A majority of teachers felt that the program provided them with the instructional background necessary to teach phonics and overall met their needs for both spelling and phonics instruction. In general, they also felt that the program was helpful in monitoring student progress. Generally, treatment students also liked the program. More than 70% indicated that they enjoyed the board work, and that the Wall Cards and letter/sound cards were helpful to them in learning and remembering phonics rules. The lowest rated item among students was the decodable readers.

When asked what they felt were the greatest strengths of the *Saxon Phonics and Spelling* program, teachers' most often cited the Wall Cards, Review Decks and daily worksheets. Teachers also noted the sight word practice, the explicit phonics instruction, and the incremental approach to teaching phonics rules (i.e., building on student's knowledge throughout the school year) as being very beneficial to student performance. In terms of other programmatic feedback provided by teachers, they noted the following: (1) pacing of the lessons was at times unrealistic given the amount of material they had to cover; (2) the amount of time spent on coding words was sometimes overwhelming to students, caused confusion for students, and at times high-level students lost interest in the lesson; (3) a few noted that it took awhile for students (and themselves) to learn the program and become accustomed to the terminology, concepts, and activities employed by the program; and (4) some teachers also noted that the program did not integrate well with their basal reading and language arts program.

In summary, this RCT with its use of quantitative and qualitative methods enabled PRES Associates to determine that the *Saxon Phonics and Spelling* program *did* produce more positive outcomes relative to classrooms that did not use this program and was associated with improved performance of students. Students who used this program outperformed students that did not in the areas of spelling, phonics, and reading words. Given the limited amount of time that this supplemental program requires (4-5 hours/week) and other factors that may have diminished differences observed, these positive effects are even more noteworthy. Moreover, results suggest that this program can help improve upon teacher's preparation and engagement in effective literacy practices. Still, further research is needed to build upon the findings from this study.

TABLE OF CONTENTS

TABLE OF FIGURES AND TABLES

Figures

Tables

PROJECT BACKGROUND

> *"The research indicates that students who learn phonics do better in all aspects of reading - word identification, accuracy of oral reading, and silent reading comprehension and fluency - than those who do not learn it. This is also true of spelling." (Chall & Popp, 1996, p. 1).*

The development of strong reading skills is essential for children to flourish in their future educational and career endeavors. Indeed, research suggests that, by 3rd grade, one can predict with a fair degree of reliability which students will ultimately drop out and which will complete their schooling based on reading skills (Slavin, Karweit & Madden, 1989). Thus, it is imperative that students increasingly develop and build on early literacy skills so as to ensure future success in the academic and occupational arenas.

According to the National Reading Panel (2000), for students to read well, they need explicit, systematic instruction in five essential areas, including: 1) phonemic awareness; 2) phonics; 3) fluency; 4) vocabulary; and 5) text comprehension. While the relative emphasis placed on these different areas may vary depending on the developmental level of the child, comprehensive reading programs must address all five areas – and it is important that they do so early on (Slavin, 1989). It is particularly important to emphasize phonics and phonemic awareness at the early primary level (K-2).

> *"Reading involves a wide variety of skills, all of which impact a student's ability to derive meaning from text. Two of these skills are: (1) the ability to understand the relationship between spoken language and sounds (phonemic awareness), and (2) the ability to translate written symbols to sounds (phonics). These skills have a substantial impact on students' initial success in reading and, consequently, on their later success to learn effectively from text." (National Reading Panel, 2000, para. 5).*

In order to help provide students' with the skills they need to become successful readers, Saxon Publishers released *Saxon Phonics and Spelling*, a K-2 program designed to supplement existing classroom reading programs. Through incremental development of new skills and continual review throughout the year, *Saxon Phonics and Spelling* aims to provide students with confidence as well as a strong foundation for independent reading. In order to accomplish its goals, a structured, systematic, and multi-sensory program was created that integrates explicit and systematic instruction of the foundational skills of phonics, spelling and reading.

Planning, Research, and Evaluation Services (PRES) and Associates[1] conducted a year-long study to examine the effectiveness of the *Saxon Phonics and Spelling* supplemental program among students in 1st and 2nd grades. This randomized control trial (RCT), which commenced in the Fall of 2006, was designed to fully address the quality criteria

[1]PRES Associates is an external, independent, educational research firm with more than 15 years of experience in applied educational research and evaluation. For more information, please visit www.presassociates.com.

put forth by the What Works Clearinghouse (WWC) in the Study Review Standards[2].

PROJECT OVERVIEW

The overarching purpose of this RCT was to rigorously evaluate the effectiveness of the *Saxon Phonics and Spelling* program in helping elementary school students attain critical reading skills. Specifically, the study was designed to address the following key evaluation questions:

1. Do phonics, reading words, and spelling improve over the course of participating in the *Saxon Phonics and Spelling* program? Does this vary across different types of students and levels of implementation?

2. How does phonics, reading words, and spelling performance differ between students who use *Saxon Phonics and Spelling* as compared to students who do not use this program? Do effects on student achievement differ across types of students or settings?

3. Does participation in *Saxon Phonics and Spelling* result in other positive student outcomes (e.g., positive attitudes towards reading and so forth)?

4. What did the users of *Saxon Phonics and Spelling* think of the program?

The remainder of this report includes: 1) a description of the design and methodology; 2) sample and site information, including descriptions of *Saxon Phonics and Spelling* program implementation; 3) results of the evaluation; and 4) conclusions. In addition, Technical Appendix E (pages 96-122) presents details of all baseline, attrition, power, and assessment analyses, including the analytical goals and framework employed.

DESIGN & METHODOLOGY

Research Design

Given the practical and fiscal constraints associated with conducting research in applied educational settings, researchers designed this study in such a way as to maximize the potential of this study in meeting all standards and criteria described in the WWC Study Review Standards. Appendix A outlines how this study addresses and/or meets each of the WWC Study Review standards.

The research design consisted of a one-year randomized control trial, with random assignment of teachers to a treatment (i.e., use of *Saxon Phonics and Spelling*) or control group (i.e., no use of *Saxon Phonics and Spelling*) *within* schools[3]. Other important design and methodological features include:

[2] A copy of this document can be obtained online at http://www.whatworks.ed.gov/reviewprocess/study_standards_final.pdf.

[3] There are a number of reasons that PRES Associates chose assignment to treatment conditions be done at the classroom level *within* schools. The most important reason for selecting this level of assignment is that such a design helps to establish *causality* by eliminating the threat that school level factors could have potentially contributed to differences between treatment and control groups. For instance, a school might have had a condition (besides the treatment) that may influence student performance on the outcome measures. Since treatment and control groups were within the same school, school level explanations of differences were reduced. An important issue to be considered with this design option, however, is that procedures must be put into place to ensure that the treatment and control groups are not contaminated through teacher contact with one another. Indeed, this was accomplished through stringent guidelines provided to the teachers and close monitoring of their use of resources.

- The study was conducted in the 1st and 2nd grades[4].
- Clear site selection criteria were established along with accompanying rationale.
- Extensive background data were collected on instructional activities and materials employed in both treatment and control classrooms so that distinctive pedagogical elements could be described given the common content taught in their reading/language arts time.
- The threat of differential attrition was addressed via: 1) the initial site selection process[5]; 2) random assignment within schools, at the teacher level, to help ensure that attrition is relatively constant across both treatment and control groups; and 3) the characteristics of students who dropped out were statistically compared between treatment and control groups;
- Extensive implementation guidelines and monitoring procedures were embedded to measure the fidelity of treatment implementation;
- A battery of assessments, including a norm-referenced standardized assessment, measuring core reading and spelling skills were used in order to enhance the sensitivity of the study to picking up treatment effects;
- The study employed the use of statistical controls[6] as well as random assignment to establish initial group equivalence;
- The study employed pre/post measures of, among other things, (1) student performance; (2) school, teacher and reading-related attitudes; (3) teacher practices; and (4) teacher knowledge and characteristics.
- Student assessments, surveys, and classroom observation forms are valid and reliable as shown by technical documentation and statistical analyses performed.
- Analyses of assessment data were primarily conducted via Hierarchical Linear Modeling with student, teacher/class, and school level data to take into account dependency issues.

The following figure displays the timeline for the important study activities. More detailed information on these activities, as well as the measures, characteristics of the core reading and language arts curricula, sample, and sites are discussed in the following sections.

[4] The reason for including these two grade levels is that the structure and emphasis of the *Saxon Phonics and Spelling* program changes substantially from first to second grade, thus it is important to include both grade levels so that researchers can look at the effects of the program cumulatively as well as by grade. Kindergarten was not included due to variation in the structure of the school days and weeks that is typically found across the country (e.g., ½ day Kindergarten, Full-day, etc.). Such variation makes it difficult to attribute observed differences to the presence/absence of a supplemental program and not to differences in the school calendar. This combined with: a) the inherent difficulties associated with assessing students at the earliest grade levels; b) relatively minor programmatic differences between Kindergarten and 1st grade; and c) power and design sensitivity concerns; all contributed to the decision to not include Kindergarten in this RCT.

[5] Sites that historically had high student attrition were not used in the study.

[6] Random assignment helps to create group equivalence. However, it must be noted that with small sample sizes random assignment in and of itself does not assure initial group equivalence (Lipsey, 1990).

Table 1. Timeline of Activities

Activities*	September	October	November	December	January-March	April	May	June
Training and Program Implementation Begins	X	X						
Assessments and Surveys Administered	X	X					X	X
Site Observations			X	X		X	X	
Teacher Logs**		X	X	X	X	X	X	

*Sites C, D and E were late additions to the study and treatment teachers began implementing the *Saxon Phonics and Spelling* program the first week of November. Training and administration of pre-assessments occurred late October.
**Note that treatment and control teachers completed monthly teacher logs that monitored instructional activities and the use of program and other resources.

Measures

This section reviews the measures that were administered, including descriptions of the items, and available reliability and validity information.

ASSESSMENTS

In order to enhance the sensitivity of the RCT to detect any effects associated with *Saxon Phonics and Spelling*, a battery of outcome measures were selected. Assessment selection was based a thorough literature review of existing assessments to identify tests that were valid, reliable, and measured core reading and spelling skills. The skills that are targeted by the *Saxon Phonics and Spelling* program, i.e., phonics, spelling, and reading words, were also considered. Student assessments were selected to measure these constructs and consisted of the following:

- Iowa Test of Basic Skills (ITBS), a norm-referenced standardized assessment with subtests focused on measuring phonics skills via word analysis, and reading words (at the first grade level) or spelling (at the second grade level);
- Dolch Word test containing high frequency and sight words that 1st and 2nd graders are expected to read independently;
- Ganske Developmental Spelling test (1994) designed to measure the developmental spelling skills of elementary and middle school students;
- Morris and Perney (1984) spelling test developed to measure the spelling levels of K-1st grade students. This test was administered to 1st graders only.

Iowa Test of Basic Skills (ITBS) - The ITBS, Form A, published by Riverside Publishing, is a group-administered test that measures various reading skills, including phonics, reading, and spelling. At the first grade level, the word analysis and reading words subtests were selected for administration. At the second grade level, the word analysis and spelling subtests were selected for administration. The ITBS provides scale scores and percentile ranges for each subtest.

- *Word Analysis:* The 35 item word analysis subtest assesses students' phonological awareness and understanding of word parts. At the first grade level, the focus is on letter identification and letter-sound relationships. Items consist of letters, pictures, and words. The second grade level test also includes basic letter-sound questions, but more complex word-building tasks involving affixes and compound words are introduced as well.

- *Reading Words:* This 29 item subtest, administered to first graders only, measures students' ability to

read words in isolation and to use context and picture cues for word identification.

- *Spelling:* The 23 item spelling subtest, administered to second graders only, consists of students looking for a mistake in spelling among four items.

The ITBS has demonstrated validity (content, criterion, and construct-related validity). Split-half reliability coefficients[7] range from .80 to .89 for the subtests, supporting the stability of the measures. This information is described in detail in the publisher's technical manual. Furthermore, the test publisher claims that the tests are sensitive enough to measure the range of reading abilities evident in classrooms across the country.

Dolch Word List– The Dolch Word List consists of 220 high frequency words (preschool through 3rd grade). The Dolch word list is made up of "service words" (pronouns, adjectives, adverbs, prepositions, conjunctions, and verbs). A number of the words cannot be sounded out because they do not follow decoding rules and, therefore, must be learned as sight words. Although originally published in 1948, the Dolch List has held up over time as a reliable high frequency word list that is used in beginning reading programs.

For the current study, 41 and 46 words obtained from the 1st and 2nd grade lists, respectively, were used. Students were asked to read each word aloud and teachers graded their verbal response as correct or incorrect. For analytical purposes, percent correct at pre and post-testing was used.

Morris and Perney Spelling Test – The purpose of this test is to analyze students' strategies in spelling. It is also an indicator of phonemic awareness. Originally tested with students in the 1st grade (Morris & Perney, 1984), this test intentionally includes unfamiliar words to obtain an assessment of students spelling attempts.

The 12 spelling items are scored via a rubric[8]. Each spelling attempt is assigned a score of 0-4 based on the identified spelling stage. Stages include the following:

0 = random letters
1 = pre-phonetic (beginning consonant only)
2 = semi-phonetic (beginning and ending consonants)
3 = phonetic (represents consonant and vowel segments)
4 = transitional (words are represented by more conventional patterns)
5 = correct

As shown, the focus is not necessarily on the correctness of the spelling but rather on students' spelling attempts and the developmental spelling stage that the student exhibits. This measure has been found to be a significant predictor of reading achievement; correlation to Metropolitan Achievement Test's reading score was found to be .61 (Morris & Perney, 1984).

Since this assessment was designed for students in grades K-1, this test was only administered to study participants in the 1st grade. The total score (sum of all the items) was calculated for each student. Based on this total score, each student was assigned an overall spelling stage. For analytical purposes, the obtained total was divided by

[7] These reliability coefficients are based on the Kuder-Richardson Formula 20.

[8] A single researcher scored all of these tests. A sample of tests were graded first and then again at the end of the scoring period. Correlations in scores for this sample were very high (.97).

the total possible points (60). In addition, a spelling stage was identified for each student based on their overall performance.

Ganske Developmental Spelling Test– This test (1994) was created as a measure of student's developmental spelling stage. Five stages of spelling are measured: 1) preliterate stage – encompasses the writing attempts of children who do not yet read, 2) letter-name strategy – students' use knowledge of alphabet and beginning attempts at reading to attempt spelling using letter-name correspondences; 3) within-word pattern – students have a broader understanding of English orthography and use more conventional spelling; 4) syllable structure – students demonstrate an understanding of polysyllabic words, and 5) derivational constancy – students are able to make links between orthography and meaning. For more information on these stages, the reader is referred to Ganske's dissertation report (1994).

Research conducted by Ganske (1994), showed that the majority of 1st graders are in the preliterate or letter-name stages and the majority of 2nd graders are in the letter-name or within-word stages. Although most 1st and 2nd graders fall into these stages, this assessment was selected to obtain a measure that would be sensitive to more sophisticated spelling stages and would provide information on spelling stages of second graders[9].

For the current study, a modified version of Ganske's screening inventory was used. Because this assessment was created for students in grades 1-8th, modifications were made to the screening inventory so that more words encompassing the letter-name, within-word, and syllable structure stages were included. For the 1st grade version, 20 items were used from the letter-name and within-word spelling lists. For the 2nd grade version, 25 items were included from the letter-name, within-word, and syllable structure spelling lists[10]. Each spelling word was scored as correct or incorrect.

According to Ganke's dissertation research (1994), alpha values for the screening inventory was calculated at .91 (grades 1-4) and test-retest reliability was .94 (grades 1-4). The test also has demonstrated construct, content, and criterion validity (see Ganske, 1994).

SURVEYS

Student Surveys. In an effort to examine other potential areas that may be influenced by the *Saxon Phonics and Spelling* program, a student survey was developed primarily to measure:

- Enjoyment of Phonics-Related Activities, Reading, Spelling and Writing *(e.g. I like to read)*
- Perceived Reading Ability *(e.g. I am a good reader)*
- Perceived Effort and Motivation in Class/School (*e.g., I try hard at school*)
- Support from caregivers (e.g., *My parents help my with my homework*).

These scales were included in order to obtain measures of the impact of the *Saxon Phonics and Spelling* program on affective student outcomes, as well as to examine differences among other factors that may need to be controlled for between groups

[9] This is because the ITBS spelling test does not provide this information and 2nd graders were excluded from the Morris & Perney test since it was not designed for this level.

[10] It should be noted that during pretesting, the spelling lists included items from the derivational constancy stage. However, none of the students tested obtained any of these items correct and teachers noted that this portion of the spelling test was extremely difficult and frustrating for students. Therefore, these items were removed at post-testing. In addition, due to the modifications made to the original feature inventory, it was difficult to assign a spelling stage to each student with confidence.

(e.g., parental support). While some items were created by PRES Associates, others were derived from three scales with published reliability and validity. Internal consistency of the scales measuring attitudinal constructs range from .53 to .79; however only one construct (motivation) is below the .70 typical acceptability criteria. High scores represent a positive attitude or agreement (scales are from 1 to 3).

Teacher Surveys. Information was collected via surveys from all participating teachers. In addition to obtaining teacher background and demographic information which could subsequently be used as covariates, the survey was developed to measure:

- Current and past classroom practices
- Teacher knowledge and preparation for providing instruction in the five elements of reading, spelling, and writing
- Organizational factors/context
- Attitudes about student learning, literacy instruction, and their classes
- Attitudes about their curriculum
- Teacher characteristics

Some items were obtained from existing scales, while others were developed for the study. Internal consistency of the scales measuring attitudinal constructs range from .79 to .91. High scores represent a very positive attitude or strong agreement (scales are from 1 to 5).

Classroom Observation Forms. A classroom observation form was developed to guide observations. This form is largely based from existing protocols that have been used extensively and across the nation[11]. Modifications were made to reflect content and practices typical of 1st and 2nd grade reading and language arts classes, as well as to examine implementation of key components of the *Saxon Phonics and Spelling* program. A pilot study showed that the inter-rater agreement among PRES-trained observers for the qualitative and quantitative portions of the observation form was high (86%).

Procedures

To ensure that all treatment teachers participating in the study had sufficient knowledge and skills to successfully implement this program right from the start, teachers were given implementation guidelines and provided training prior to implementation. In addition, monitoring procedures (via teacher logs and classroom observations) were developed to measure the extent to which treatment teachers were implementing a similar instructional model. The following section presents the procedures used to assist treatment teachers in implementing the *Saxon Phonics and Spelling* program, the monitoring procedures used by evaluators to determine treatment fidelity, methods used to obtain program feedback, test administration, and scoring procedures employed.

TRAINING

Treatment teachers at all sites met with the author of the *Saxon Phonics and Spelling* program for a professional training session for approximately 3 hours prior to implementation of the program in their classes. For several sites, the author was joined by another professional trainer of the *Saxon Phonics and Spelling* program. This trainer conducted the follow-up training sessions with the sites. During the initial session, the author described the philosophy and key components of the program, reviewed the program materials, and worked with teachers to determine how *Saxon*

[11] The Classroom Observation Form was derived from a form constructed and validated in other PRES elementary evaluation studies. For more information on the validity and reliability of this form, please contact PRES Associates.

Phonics and Spelling lessons should be structured given their core reading curriculum. Representatives of PRES Associates attended these trainings. It should be noted that training did not focus solely on professional development (e.g. effective teaching strategies), although suggestions were offered, but rather on what the vision of the program was and how to use the *Saxon Phonics and Spelling* program to help students gain greater phonics, spelling, and reading skills. Still, teachers felt that the training was highly useful.

> *"Everyone should have the training with this program. It makes a world of difference. We had this program years ago but never had the training and it makes all the difference."*
>
> *--Saxon Phonics and Spelling Teacher*

At least one follow-up training session was conducted at each participating school. These sessions lasted approximately 2 hours and offered an opportunity for teachers to voice concerns, ask questions and get more instruction on implementing the program in their classrooms. While third training sessions were offered to all schools, only one participating site, School F, elected to have this follow-up training session to ask questions and get help with pacing of the program. Treatment teachers at School F also received additional training through observation and collaboration with teachers in a neighboring school who were long-time users of the program. PRES Associates and Saxon Publishers, arranged for treatment teachers at School F to observe these teachers, meet to ask questions and exchange information for future questions. These teachers noted that the opportunity to observe other teachers implementing model lessons within real-world classrooms, greatly assisted them in their implementation of the program. As one teacher noted:

> *"Before teachers even open their materials, they should see the program in use first. If the trainer would have come after we had observed the program, it would have been better."*
>
> *--Saxon Phonics and Spelling Teacher*

It is recommended that future research and possibly future product trainings employ visits to classrooms that are using the intervention so that study teachers can observe model lessons. This was perceived as very beneficial to study participants who had this experience.

IMPLEMENTATION GUIDELINES

Through guidance and consultation with *Saxon* Publishers, key components of the *Saxon Phonics and Spelling* program were identified. It should be noted that these key components are representative of what would typically be done in classrooms using this program. Based on this information, implementation guidelines were developed for use by the treatment teachers prior to implementation of the program (see Appendix D for a copy of the Implementation Guidelines). Teachers were instructed to "follow these guidelines as you implement the *Saxon Phonics and Spelling* program …all of these checked items are considered critical to the success of the program." The following thirteen key program components were provided to treatment teachers as elements that they should employ in their classrooms:

1. Introduce lesson (part of warm-up)
2. Engage in language-alphabet activity (part of warm-up)

3. Review letters, sounds and/or spelling (part of warm-up)
4. Teach new increment
5. Use *Saxon Phonics and Spelling* keywords, Deck Cards, and related Wall Cards (part of new increment)
6. Use student spelling dictionary and reference booklets (part of new increment)
7. Complete board work
8. Assign worksheet for students to complete
9. Conduct classroom practice (Kid Cards, independent reading, decoding/fluency activities)
10. Conduct oral and written assessment (once per week)
11. Conduct site word evaluation (once per week)
12. Devote at least 60 minutes for 1st grade and 50 minutes for 2nd grade to *Saxon Phonics and Spelling* activities/day
13. Follow *Saxon Phonics and Spelling* Pacing Guide (1 lesson per day)

For a full description of these key components, please see Appendix D.

Program Monitoring

Teacher Logs. Online teacher logs were developed so that program implementation could be monitored on a real-time basis. Teachers were instructed to complete these on a monthly basis from October (or November for late addition schools C, D, and E) through May. Both treatment and control teachers completed teacher logs, with slightly different versions for each. The primary purpose of the teacher logs was to monitor program fidelity of the treatment teachers. The reason researchers also collected monthly logs from control teachers was so instructional activities and content covered could be monitored. Such background information provided researchers with a detailed data source on what was occurring in treatment and control classrooms in terms of language arts and reading instruction. This allowed researchers to identify areas of overlap in terms of content taught and activities in treatment and control classrooms. The extent to which there are similarities and differences between classrooms can have an impact on observed differences between treatment and control classes and effect sizes. Thus, it is important to take these factors into consideration. Information obtained via these logs included changes in their student roster, typical reading/language arts classroom activities, use of other resources and exercises (including homework and independent practice), and for treatment teachers, use of key *Saxon Phonics and Spelling* program components.

Results showed that teachers had, on average, a 69% completion rate. The ranges were 20% to 100%[12]. Teachers were contacted after failure to complete teacher logs each month. In cases of noncompliance, the school liaison was asked to confer with the teacher to see if there was anything that could be done to assist the teacher in completing the logs. Furthermore, for those teachers that did not have high completion rates, a more extensive implementation checklist and interview was completed during the Spring site visit to ensure that information on implementation, instructional practices, and classroom activities was available.

Classroom Observation. Classroom observations were conducted for all treatment and control teachers during the Fall/Winter (October to January[13], 2006-7) and the Spring (April to May, 2007). The purpose of these observations was to better understand the instructional approaches and

[12] Calculation based on 7 months in which teachers were asked to report on their activities.
[13] While a formal observation took place in January for Site C, a phone conference to monitor implementation took place in December.

materials used by teachers with their students and to identify differences and similarities between teachers who were randomly assigned to treatment and control conditions. Specifically, information was obtained on how reading, language arts and *Saxon Phonics and Spelling* class activities were structured, what and how materials were used, and characteristics of the class including student engagement, classroom environment and culture, and teacher-student interactions. In addition, teachers were interviewed after the observations to obtain more specific information on the representativeness of the lesson, resources used, ability levels of the students, assessment practices, pacing, independent practices, and test preparation strategies. The observations also allowed researchers to examine the extent to which teacher level differences could have influenced study results and to examine possible contamination between treatment and control teachers.

TEST/SURVEY ADMINISTRATION AND SCORING

All assessments and surveys were administered during the Fall (September through October, 2006) and Spring (May through June, 2007)[14]. The tests and surveys were administered by teachers. A standard testing procedure, based on the publishers'/authors' administration instructions, was distributed to all teachers prior to testing. Teachers were instructed to contact PRES Associates if they needed further guidance. The ITBS subtests were scored by PRES researchers following the standardized scoring procedures (including raw score conversions) as outlined in the ITBS Scoring Manual, without regard to group assignment. As previously noted, spelling tests were scored by guidelines set forth by authors (Ganske, 1994; Morris & Perney, 1984).

[14] Administration dates depended on the school's start and end date. Teachers within each school followed a similar testing schedule. With the exception of sites C, D, and E, administration occurred within 6 weeks after the school year commenced (pretest) and within 1 month prior to the end of the school year (posttest).

Curricula

SAXON PHONICS AND SPELLING

The *Saxon Phonics and Spelling* program is an intervention designed to supplement existing classroom reading programs. *Saxon Phonics and Spelling* focuses on phonemic awareness, alphabetizing, decoding, spelling, and increasing reading fluency in elementary school students. This program sets itself apart from other basal readers because it uniquely integrates explicit and systematic instruction of the foundational skills of phonics, spelling and reading. The program emphasizes 7 key elements including:

- Controlled vocabulary and reading practice
- Role of literature
- Coding
- Spelling
- Assessment and remediation
- Handwriting
- Reading fluency

To accomplish the goals of the program, the entire *Saxon Phonics and Spelling* program includes 100% controlled reading, where new increments of learning are introduced each day and always builds on previously taught concepts for continual review. The controlled practice is built in so that students can always retrieve the information they need. In addition, a multi-sensory approach (combination of visual, auditory, and kinesthetic elements) is emphasized during instruction.

As noted in Appendix D, the program includes a variety of components to meet the

instructional needs of teachers and students. The Teacher's Manual consists of binders that include a scope and sequence, handwriting instructions for the alphabet, spelling rules, listing of support materials, and removable lesson booklets for the daily whole group lesson. The Teacher Resource Binder includes home letters (so parents can assist their children in reviewing skills taught in class), masters, and recording forms for assessments. Teachers are also provided with review decks for continual review of previously taught concepts and skills. Review decks consist of letters, pictures, spelling, sight words, affixes, and alphabet/accents. Wall Cards, Kids Cards, and posters serve as reference and review tools for the class. Students are provided with worksheets, decodable and leveled fluency readers, alphabet handwriting strips, letter tiles, and a spelling dictionary and reference booklet.

The *Saxon Phonics and Spelling* lesson is comprised of three parts: (1) warm-up, (2) new increment, and (3) application and continual review. More information on how these lessons were conducted in the participating classrooms is provided in the case studies in Appendix B and on page 59. In addition, for a detailed description of the program's key features and materials, see Appendix D-Implementation Guidelines. Treatment teachers were provided with all components of the *Saxon Phonics and Spelling* program as noted in Appendix D[15].

CORE READING/LANGUAGE ARTS CURRICULA

It should be noted that the *Saxon Phonics and Spelling* program is designed to be used as a supplemental program with *any* core reading/language arts program. Similarly, in this study the *Saxon Phonics and Spelling* program was used in conjunction with a variety of core curricula. For detailed information on each reading and language arts curriculum, the reader is referred to the case studies in Appendix B. The following presents a summary and comparison of the various core reading and language arts curricula used in the participating classrooms.

There were a total of five distinct basal programs used by the schools participating in the study. Schools D and E used the same core reading program (A[16]) published in 2001. Schools A and C used the same basal program (B), also published in 2001. School F used a 2003 basal program (C). Because school B divided students for reading instruction based on ability levels, two different programs were used. Two teachers (one control and one treatment) at school B used the same program (D), published in 2002, while the third teacher (treatment) used an older program (E), published in 1993. Therefore, with the exception of school B, both treatment and control teachers within each school had the same basal program at their disposal.

In general, the curricula are comprehensive, basal reading and language arts programs, designed to be used as the main resource for these subjects. Indeed, for the most part, teachers used their respective program as their main resource for reading and language arts instruction. Exceptions to this included the teachers at school E who supplemented their program extensively, engaged in guided reading, and used their district guidelines for lesson planning. These teachers then used whatever resource they had available, including basal program resources, to meet their teaching needs. The

[15] Appendix C contains tables on the percentage of use of the *Saxon Phonics and Spelling* resources.

[16] To ensure confidentiality of the school and basal program, the name of the program employed is excluded. However, descriptions of each program are provided in the case studies in Appendix B.

control teacher at school B and the 1st grade treatment teacher at school A also tended to supplement the basal program extensively. All other teachers noted using other supplemental materials, including some devoted to phonics and spelling instruction, though they tended to use their basal program as the main source for instruction. Furthermore, teachers at all participating sites followed a similar set of standards prescribed by the district or school and as such, they were teaching very similar content, although at times used different materials to do so.

With the exception of school B, where a much older program was employed (1993), most of the core reading/language arts curricula used at the sites were published within just a few years of one another (2001-2002) and all shared similar features. Similarities among the reading and language arts curricula included the following:

- Use of authentic literature to teach reading, vocabulary, comprehension, writing, spelling, and other language skills.
- Opportunities for multiple readings of literature.
- A weekly five-day lesson sequence using a literature passage for each new lesson.
- Oral discussion and practice opportunities prior to independent practice.

Generally, the content included in each of the various curricula was similar; however, there were some notable differences. The curriculum used at schools A and C provided more in-depth instruction on phonics as well as a greater focus on grammar and mechanics than the curricula at the other sites. The curriculum used at sites D and E, conversely, had a much more limited focus on phonics and mechanics. Additionally, the two curricula employed at site B placed a greater emphasis on independent student reading through incorporation of leveled readers or self-selected independent reading selections. Additionally, the authentic literature passages used in the different programs were unique to each program and offered different opportunities for instruction.

In sum, within school random assignment helped control for differences in curricula use. The vast majority of treatment and control teachers within each school used the identical core reading/language arts curriculum. As noted in the Technical Appendix, there were also no significant differences in the time spent on reading, writing, spelling, and phonics during their core reading and language arts instruction between treatment and control teachers. Overall, monitoring of instructional practices showed that with minor variation, the main difference between control and treatment classrooms was the presence or absence of *Saxon Phonics and Spelling*.

With the exception of school B, both treatment and control teachers within each school had the same core reading/ language arts curriculum.

Site Selection Criteria

Sites were selected[17] using the following criteria:

- Public school;
- Relatively low % of second-language

[17] In addition to obtaining referrals from Saxon representatives and using existing contacts, PRES Associates obtained a list of schools that met key selection criteria (e.g., relatively low attrition rates, diverse student ethnic backgrounds, etc) from a market research firm. This company has a comprehensive database of virtually all schools across the country. Districts/schools were randomly selected from the lists and contacted until a sufficient sample agreed to participate in the study.

learners (e.g., less than 33%);
- Located in Continental U.S.;
- Willingness/commitment to fully participate in all aspects of the study

Other major criteria taken into consideration included: 1) reading instructional time available; 2) the school has relatively high attendance and historically low student mobility rates; and 3) diversity in geographical location.

SAMPLE DESCRIPTION

Site Characteristics

A total of 6 elementary schools participated in the *Saxon Phonics and Spelling* RCT. Figure 1 displays the geographical location of each of the participating sites. As shown, there is a good variation in the geographical location of the sites and there is a mix of urban, suburban, and rural schools.

A detailed case study of sites is provided in Appendix B. It should be noted that based upon historical state assessment data and as discussed in the case studies, ability levels within the student populations were fairly consistent among sites. Teachers reported that overall the classes included in the study contained a broad-range of abilities, with some variance between classes. Furthermore, monitoring of sites showed no evidence of a local history event or disruption.

Table 2 on the following page shows characteristics of each of the participating sites. The sample includes a lower proportion of Whites and a higher percentage of African Americans as compared to those found nationwide. In addition, this sample included a higher proportion of students receiving free/reduced lunch as compared to the national norm.

Figure 1. ***Saxon Phonics and Spelling*** **RCT Study Sites**

Table 2. Final Study Sample: Characteristics of Sites Participating in the *Saxon Phonics and Spelling* RCT *(2006-2007)*

Elementary School	05-06 School Size	Avg. Class Size	Teachers/ Grades (tx/ct)	Total Participating Students	Ethnic Distribution	Special Education %	English as a Second Language %	Free/ Reduced Lunch %	Gender Distribution
School A Oklahoma	377	26	4 Total G1=2 (1/1) G2=2 (1/1)	91 Total G1=47 G2=44	White-25.3% Hispanic-58.2% African Am-8.8% Nat. Am.-6.6% Other-0.6%	4.4%	46.5%	NA *83%* *school-wide*	Male – 49.5% Female – 50.5%
School B Idaho	610	22	3 Total G2=3 (2/1)	65 Total G2=65	White-92.3% Hispanic-3.1% African Am-3.1%	4.6%	10.2%	58.5%	Male – 41.5% Female – 58.5%
School C Indiana	515	22	8 Total G1=4 (2/2) G2=4 (2/2)	175 Total G1=91 G2=84	White-97.1% African Am-0.6% Asian-1.7%	29.1%	3.5%	*26.9%*	Male – 54.3% Female – 45.7%
School D Texas	680	22	4 Total G1=2 (1/1) G2=2 (1/1)	77 Total G1=44 G2=33	White-14.3% Hispanic-85.7%	NA *9%* *school-wide*	85.5%	14.5%	Male – 58.4% Female – 41.6%
School E Texas	568	19	6 Total G1=3 (2/1) G2=3 (2/1)	92 Total G1=58 G2=34	White-5.4% Hispanic-51.1% African Am-43.5%	NA *9%* *school-wide*	29.5%	NA *92%* *school-wide*	Male – 51.1% Female – 48.9%
School F Georgia	563	19	10 Total G1=5 (2/3) G2=5 (2/3)	182 Total G1=95 G2=87	White-3.8% Hispanic-8.2% African Am-81.3% Asian-1.1% Other-5.5%	12.6%	15.0%	83.0%	Male – 50.0% Female – 50.0%
Overall (avg.) ***(Across all sites)***	552	22	35 Total G1=16 G2=19 TX=18 CT=17	682 Total G1=335 G2=347 TX=312 CT=370	White-40.5% Hispanic-26.8% African Am-29.2% Asian-.9% Nat. Am.-1.0% Other-1.6%	13.7%	18.8%	60.4%	Male – 51.3% Female – 48.7%
National Population					**White-55.9%** **Hispanic-20.5%** **Black-16.9%** **Asian-4.5%** **Other-1.2%**	**13.6%**		**40.9%**	**Male-50.9%** **Female-48.1%**

NA=Not available at the student level.

Student Characteristics

The final sample consisted of 682 students (49.1% in 1st grade and 50.9% in 2nd grade). Table 3 presents the demographic distribution among study participants. Note that only students who remained in the study throughout the year are included in this table.

Preliminary analyses[19] were performed to examine whether baseline differences existed as a function of student demographics. Chi-square analyses on the demographic characteristics noted in Table 3 showed that two of the eight comparisons were significantly associated with group,

Table 3. Student Demographics Distributions*

Characteristics		Control (n=312)		Treatment (n=370)		Total (n=682)		National[18]
		Count	Percent	Count	Percent	Count	Percent	Percent
Gender *($\chi^2(1)=0.11$, $p=.75$)*	**Male**	158	50.6%	192	51.9%	350	51.3%	50.9%
	Female	154	49.4%	178	48.1%	332	48.7%	48.1%
Ethnicity *($\chi^2(5)=2.16$, $p=.83$)*	**White**	121	38.8%	155	41.9%	276	40.5%	55.9%
	Hispanic	83	26.6%	100	27.0%	183	26.8%	20.5%
	African Am.	95	30.4%	104	28.1%	199	29.2%	16.9%
	Asian	4	1.3%	2	.5%	6	.9%	4.5%
	Native Am.	3	1.0%	4	1.1%	7	1.0%	1.2%
	Other	6	1.9%	5	1.4%	11	1.6%	--
Grade *($\chi^2(1)=4.47$, $p=.04$)*	**1st**	167	53.5%	168	45.4%	335	49.1%	--
	2nd	145	46.5%	202	54.6%	347	50.9%	--
Subpopulations								
($\chi^2(1)=.02$, $p=.88$)	**English as Second Language****	55	18.6%	67	19.0%	122	18.8%	--
($\chi^2(1)=1.77$, $p=.18$)	**Special Education**	44	15.7%	37	11.9%	81	13.7%	13.6%
($\chi^2(1)=1.08$, $p=.30$)	**Free/Reduced Lunch**	161	62.6%	140	58.1%	301	60.4%	40.9%
($\chi^2(1)=0.001$, $p=.98$)	**Low Literacy Level**	60	19.7%	72	19.8%	132	19.8%	--
($\chi^2(1)=3.84$, $p=.05$)	**High Literacy Level**	59	19.4%	50	13.8%	109	16.3%	--

* Counts (and percents) do not include missing information. Sites A and E could not provide free/reduced lunch status information for students (n=81). Site D and E could not provide information on special education status of its students. Literacy level was determined by using pretest ITBS percentile rankings (bottom 30% vs top 70%) across the two subtests administered at each grade level. Students for whom categorization was not consistent across both subtests or who had missing data on one or both tests were excluded.
Results showed only 7 students as missing more than ¼ of the school year (5 treatment and 2 control students). There was no relationship between group and attendance. This information is excluded from the table above because measure not taken at baseline.
** This measure is based on self-report. Students were asked to indicate the primary language spoken at home. Counts represent students who indicated "Spanish" or "other". This measure was found to have a high correlation with the school-reported Limited English Proficiency (LEP) category (r=.46). The self-report measure is used in the majority of analysis because of the extent of missing information on LEP status (two schools did not report).

[18] Data was obtained from SchoolDataDirect.com, an online service of the Council of Chief State School Officers' State Education Data Center. Figures represent distributions across all grade levels and reported for 2006.

[19] All details regarding analyses on baseline differences are provided in Appendix E (the technical appendix) on pages 100-101.

$_{p<.05}$[20]. Results showed that there were significantly more 2nd graders and less 1st graders in the *Saxon* group as compared to control group. In addition, there was a higher proportion of control students classified as being high-performing in reading skills as compared to *Saxon* students. However, as described in the following paragraph, no significant pretest differences were observed between *Saxon* and control students. Overall then, these results suggest that at baseline, treatment and control students' characteristics were comparable on the majority of demographic characteristics[21].

Examination of the norm-referenced ITBS pretest showed that participating students, on average, were at the 49th percentile of the ITBS norm sample. This indicates that this sample is on par with the national sample. In addition, pre-test differences between treatment and control students on the assessment measures were also examined. The student level analyses did not reveal any significant differences on all six outcomes analyzed (this includes all subtests, see Table 4). Nevertheless, pretest scores and demographics were used as covariates in multilevel models in order to enhance the analyses' power to detect treatment effects.

Table 4. Sample Size, Means, Standard Deviations, and *t*-test (Student Level) Results for Assessments at Pre-testing

Pretest*	Group	N	Mean	Std. Dev.	*t*	Sig. Level
ITBS – Word Analysis SS (1st and 2nd graders)	Control	304	149.33	20.84	0.859	.391
	Treatment	363	147.99	19.26		
ITBS – Reading Words SS (1st grade only)	Control	164	139.43	11.38	-0.283	.778
	Treatment	164	139.82	13.88		
ITBS - Spelling SS (2nd grade only)	Control	140	156.95	12.73	1.646	.101
	Treatment	199	154.83	10.91		
Morris & Perney Spelling Test – PCT (1st grade only)	Control	151	58.57	23.08	-0.530	.597
	Treatment	163	59.98	24.15		
Ganske Spelling Test – PCT (1st and 2nd graders)	Control	315	16.52	13.95	-1.218	.224
	Treatment	339	17.82	13.20		
Dolch Words – PCT (1st and 2nd graders)	Control	302	63.17	36.43	-1.873	.061
	Treatment	355	68.30	33.28		

*SS=Scale Score; PCT=Percent of total

[20] "Significant" means that we can be 95% or more confident that the observed differences are real. If the significance level is less than or equal to .05, then the differences are considered statistically significant. If this value is greater than .05, this means that any observed differences are not statistically significant and may be interpreted as inconclusive. However, at times this may be referred to as "marginally significant." In this case, the criterion is more liberal and means that we can be 90% or more confident that the observed differences are real.

[21] Note that program effect analyses control for both grade (on those assessments where both 1st and 2nd graders took the test) and on pretest scores.

Teacher Characteristics

There were 35 teachers (17 control and 18 treatment) who participated in the RCT. Thirty-one teachers are female and three are male. Teachers were predominantly Caucasian (70.6%).

In regards to educational background, 1 has a Ph.D. in Curriculum and Instruction, 8 teachers have a Master's degree in Curriculum, General Education, or Elementary Education. The remaining 26 have a Bachelor's degree in Early Childhood Education, Elementary Education, Special Education or Physical Education/English. All teachers are state certified to teach at the elementary grade level.

Teacher experience ranged among the participating teachers. Teachers had taught from 1 to 34 years, with the average number of years taught being 10. When asked about their experience in teaching at their current grade level specifically, results showed that teachers had taught between 1 to 32 years, with the average being 6 years. Four teachers were in their first year of teaching (2 treatment, 2 control).

Given the caveat that analyses at the teacher level has an inherent lack of power to detect differences due to the small sample size (n=35), preliminary analyses were run to examine whether baseline differences existed as a function of teacher and classroom characteristics. Results[22] showed no significant baseline differences among teachers in terms of degree earned or teaching experience. There were also no differences on affective measures such as perceptions of control over instructional decisions and barriers to effective teaching. However, significant differences were observed in terms of their knowledge and preparation to teach the five elements of reading, spelling, and writing, and their engagement in effective literacy practices. In particular, treatment teachers were less likely to have knowledge and preparation in these areas, and to engage in best practices. Given these significant differences, analyses comparing treatment and control groups controlled for these variables.

Classroom environment was also analyzed based on information collected during the Fall. Again, results showed no significant differences between treatment and control teachers in terms of how their classes were structured (whole group v.s. small groups v.s. individuals), availability of instructional materials, design of their lessons, teacher-student interactions, class culture, general lesson content, class engagement, and overall class climate.

In summary, although some differences were found between treatment and control teachers, randomization was reasonably successful in producing equivalent treatment and control groups in terms of teacher and classroom characteristics, and quite successful in achieving equivalence in terms of baseline student outcomes. Nevertheless, care was taken to include the variables that were non-equivalent between the treatment and control groups as covariates in the analyses of program effects.

Attrition Analysis

There was an overall attrition of 9% due to students leaving school or shifting from control to treatment classes (or vice-versa). Details on the attrition analysis is presented in Appendix E, pages 103-106. Results showed that there was no evidence for differences in dropout attrition between treatment and control students; that is, the

[22] For detailed statistics, please refer to pages 100-101 in the Technical Appendix E.

proportion and characteristics of those who dropped out of the study were comparable among both treatment and control groups. Results also suggest that measurement attrition was not problematic. While the lack of Dolch post-testing by one teacher (at site F) resulted in less control students taking the Dolch Word test as compared to *Saxon Phonics* students, there were no *performance* differences between group and those who completed tests and those that did not. Thus, overall results suggest that attrition is unlikely to bias results.

Comparison of Core Reading and Language Arts Instructional Practices Between Treatment and Control Teachers

Given that the majority of teachers within each site were using the same core reading/language arts program and following the same guidelines, treatment and control teachers taught very similar content within each site. Additionally, teachers taught similar content due to the fact that there were certain district and state scope and sequence that needed to be covered. The most obvious difference between teachers' instruction was the addition of the supplemental *Saxon Phonics and Spelling* curriculum to the treatment classes. For the most part, teachers within each site followed the core reading/language arts program to an equal degree. Exceptions to this were a control teacher at school B and a treatment teacher at school A; both supplemented the core reading program extensively while the remaining teachers tended to more closely adhere to the basal. Furthermore, teachers stayed close together in terms of pacing and content covered. The only exception was at school B; the ability grouped classes required that different content be covered to meet students various needs.

All teachers spent on average 1.5 to 2 hours, ranging from 1 hour to 2.5 hours on reading and language arts instruction per day. In general, both treatment and control teachers spent an equal amount of time focusing on different language arts topics within each lesson. Statistical analysis revealed no significant differences between treatment and control teachers in terms of time spent in the areas of reading, spelling, writing, and phonics, $_{p>.05}$, during their core reading and language arts time in the Fall and Spring. Anecdotally, treatment teachers indicated that they devoted a greater amount of time to phonics instruction during their reading/language arts time (as separate from their *Saxon Phonics and Spelling* lesson). However, since it is the nature of the *Saxon Phonics and Spelling* program to incorporate strategies into other portions of reading instruction, this is most likely a programmatic effect. That is, a greater emphasis on phonics in their core reading time is to be expected from teachers using the *Saxon Phonics and Spelling* program.

The structure of the lessons generally followed a similar pattern in treatment and control classrooms across sites as well. Generally, teachers began each lesson with some type of lesson introduction or a warm-up activity consisting of review exercises or other exercises designed to engage students for the lesson. One exception was school E; each lesson began with small groups rotating through a variety of center activities. The activity that followed the lesson warm-up/introduction varied depending on the program that was used. Typically, classes consisted of reading or re-reading the literature associated with the lesson, whole class activities to practice vocabulary and/or comprehension, or other activities such as writing or completing worksheets. Independent practice was generally provided to students via worksheet exercises or

writing prompts. Homework was typically assigned on average 3-4 nights per week. There were also no differences between treatment and control teachers in terms of the amount of time to complete and in the percent of students who typically completed and returned their homework, $p>.05$.

In terms of other common pedagogical practices (i.e., how lessons were delivered) employed by effective teachers, both treatment and control teachers noted that they:

- Used observation of students and oral questioning daily during reading/language arts
- Used high frequency word lists
- Asked students to make verbal or visual connections to new words
- Had students reread the same passages to practice fluency
- Connected previously taught skills to new content/texts
- Taught letter-sound relationships
- Formally assessed students and helped them prepare for standardized tests
- Modeled decoding or encoding strategies

However, as previously noted, significant differences were observed in terms of their knowledge and preparation to teach via the five elements of reading, and spelling and writing, $t(32)=4.087$, $p<..001$, and their engagement in effective literacy practices, $t(33)=3.459$, $p<.001$, during the Fall. In particular, results showed that at baseline, control teachers reported having more knowledge and preparation for teaching important elements of reading (fluency, phonics, phonemic awareness, comprehension and vocabulary), spelling, and writing, and engaged in effective literacy practices to a greater extent than treatment teachers. Because these variables can have an impact on student outcomes, analyses comparing treatment and control groups controlled for initial teacher knowledge, preparation, and engagement in effective literacy practices.

In sum, with the exception of the above findings, there were a number of common elements between treatment and control classrooms (in terms of language arts and reading content covered and broad pedagogical practices employed, e.g., how classes were structured). This is to be expected since both treatment and control teachers within sites were following the same standards and the majority were using the same core reading and language arts curricula.

Furthermore, it should be noted that a number of strategies and activities employed in core reading and language arts classes were also common to the *Saxon Phonics and Spelling* program. For example, use of high frequency word lists and teaching letter-sound relationships are key components of the *Saxon Phonics and Spelling* program, and were also used in a number of control classes. Similarities in the core reading/language arts programs or supplemental phonics programs and *Saxon Phonics and Spelling* can diminish program effects that may be associated with the *Saxon* program. Therefore, results obtained should be interpreted with this in mind—any positive effects obtained occurred despite some similarities in pedagogical practices employed in treatment and control classrooms.

Implementation of the Saxon Phonics and Spelling Program

The *Saxon Phonics and Spelling* program is designed to be used in the classroom daily for at least 50 minutes. Ideally, teachers would have as much as 90 minutes to devote to the program. Teachers reported that in actuality, they spent an

average of 58 minutes each day on the program, ranging from 45 minutes to 70 minutes. It is important to note that the integration of the program into each class depends on the existing curriculum and the format for reading/language arts instruction. For instance, at school B, *Saxon Phonics and Spelling* instruction took place in the homeroom classes, while students were rotated to other teachers for reading instruction as part of ability grouped classes. Despite this, the amount of total time spent on reading and language arts instruction, including *Saxon Phonics and Spelling,* was similar across the treatment and control classes. This is because homeroom control teachers also targeted reading skills while *Saxon Phonics and Spelling* was being used by the treatment teachers. In the remaining schools, a specific allotted 1 to 2.5 hour block was available for all reading and language arts instruction and treatment teachers needed to fit *Saxon Phonics and Spelling* lessons into that time. However, the amount of total time spent on reading and language arts instruction, including *Saxon Phonics and Spelling,* was significantly higher in the treatment classes as compared to the control classes, $t(33)=2.72, p=.01$. That is, while treatment teachers had to fit in *Saxon Phonics and Spelling* within their allotted reading and language arts time, they also tended to expand their instruction by approximately 24 minutes more as compared to control teachers.

The *Saxon Phonics and Spelling* program is designed to be taught in sequence, which meant that treatment teachers at all participating schools were teaching the program in the same order. However, there were several issues at the onset of the study related to pacing. Teachers expressed concerns that they had trouble fitting the required parts of each *Saxon Phonics and Spelling* lesson into 1 day. Over the course of the school year, as teachers became more comfortable with the program, they reported that this became less of an issue. Indeed, as previously noted, teachers at School F, who received additional training through meeting with and observing other teachers using *Saxon Phonics and Spelling* reported that this was extremely helpful to them in addressing how best to use the program and dealing with pacing issues.

Teachers used mostly whole group instruction but tried to offer students individual opportunities to practice the letters or sounds in the review decks. They did this by calling on a variety of students individually rather than instructing entirely via whole group. Treatment teachers began each daily lesson by introducing the lesson, including the objective. Following this, teachers continued with the lesson warm up by using the Deckcards to review letters and sounds. Most teachers noted that they shuffled the decks regularly, as instructed by the trainers. Only two teachers reported not shuffling the decks regularly, though they did note that they shuffled the decks on occasion. Several teachers also noted that while they used the language-alphabet activity, it was not used consistently.

Following the introduction and warm-up, teachers continued the lesson by teaching the new increment. They would introduce new letters, keywords or teach new letters and sounds as the lesson required. board work, consisting of students practicing the new lesson skill on the board in front of the class followed the lesson. Then the *Saxon Phonics and Spelling* daily worksheet was passed out. One side of the worksheet was assigned to students to be completed in class, individually. Teachers would then quickly check the completed side of the worksheet.

Based on observations, teachers were very consistent in implementing the board work and then assigning the worksheets. However, there was some variation in how teachers used the worksheet. While most teachers followed the implementation guidelines and assigned the first side of the worksheet as independent practice, the second side was not always assigned as homework. Rather, the teachers sometimes used it as in-class work as a whole group or independent practice. Still, it was very rare for teachers not to assign the worksheet at all.

The key component that was least likely to be used was the classroom practice section of the lesson. This included using Kid Card games, activities, reading and other practice, which typically followed the board work and worksheet. In interviews, teachers indicated that while they thought that the games, activities and other practice opportunities were worthwhile, they felt that they simply did not have time. Some teachers would try to use the classroom practice activities at a separate time from the *Saxon Phonics and Spelling* lesson, for example, during free time that became available after a math lesson or just before lunch. In terms of *Saxon Phonics and Spelling* assessment use, while teachers used the oral and written assessment regularly, they used the sight word evaluations less often. The reason given for lack of use was time constraints.

In terms of structure of the *Saxon Phonics and Spelling* lessons, teachers reported that on average 65-85% of each lesson was done using whole-group instruction. However, school A used more of each class block for independent practice or small group work as compared to other sites.

It should be noted that there were several optional components offered as part of the *Saxon Phonics and Spelling* program including handwriting instruction, Fluency and Decodable Readers, and Fluency Assessments. Most schools opted to use the handwriting program already in place at their school in order to be consistent with instruction students had received in previous grades. At school E, teachers used the *Saxon* handwriting only for remediation. The Fluency and Decodable readers were used by several schools as deemed necessary by teachers, but not on a regular basis.

Another item of note is that several treatment teachers expressed a desire to continue using their own spelling program instead of the spelling component of the *Saxon Phonics and Spelling* program. Specifically, some teachers wanted to use the spelling lists that they had from their school/district or from the basal program. While the *Saxon* spelling component (which consists of a word list that is incorporated into daily activities on a rolling 5-day schedule) is integrated into the *Saxon Phonics and Spelling* program, the trainers felt that this would be acceptable since the core spelling assistance would come from other parts of the program (e.g., the explicit phonics instruction, including coding and learning of phonics rules, etc.). It was noted that over the course of the year, one teacher at school A, two teachers at school C, one teacher at school D, the majority of teachers at school F, and both teachers at school E did not use the *Saxon Phonics and Spelling* program's spelling component to a great extent, if at all, with their students. However, as one teacher noted:

> *"We always use the rules and skills learned in Saxon Phonics to help decode and spell words."*

FIDELITY OF IMPLEMENTATION

Triangulation of the available information[23] showed that while the treatment teachers did an adequate job of completing a sufficient number of *Saxon Phonics and Spelling* lessons, implementation of all key program components as prescribed in the implementation guidelines was more problematic. Three levels of implementation (low, moderate, and high) were assigned for treatment teachers' implementation of key program components and the percent of lessons completed. Moderate to high intervention fidelity (with moderate fidelity being the norm) was evident among the majority of treatment teachers when it came to lesson completion (see Table 5). However, low implementation of the components was evident among 6 participating teachers. In particular and as previously alluded to, the components that teachers had more difficulty implementing included the language/alphabet activity during warm-up, the classroom practice activity, use of the student spelling dictionary and reference booklets during new increment, weekly sight word evaluation, and completing 50 to 60 minutes of daily *Saxon Phonics and Spelling* instruction. It should be noted that this less than ideal level of implementation can reduce positive effects that may be associated with the program. This means that any positive impacts observed will have occurred despite lower levels of implementation among some treatment teachers.

In addition, there was some slippage evident among a small number of treatment teachers. Teachers at school A decreased their use of the *Saxon Phonics and Spelling* program from 90 minutes daily to 60 minutes due to increasing needs to use time for other academic areas. In addition, these teachers did not use the Deckcards nor language/ alphabet activities with much consistency toward the end of the year because teachers felt the students "got it." Teachers at school D indicated that lesson completion was compromised toward the end of the year due to the increased level of difficulty students were experiencing with the program. These issues are important to keep in mind since they can reduce effects.

In sum, due to the variability in implementation levels[24], analyses are performed to examine if level of implementation affected the effectiveness of the *Saxon Phonics and Spelling* program; this is reported in the Results section.

Table 5. Level of *Saxon Phonics and Spelling* Implementation (N)

Level of *Saxon Phonics and Spelling* Implementation	Key Program Components Implementation	Percent of *Saxon Phonics and Spelling* lessons completed
Low	Less than 69% of goals met (i.e., 8 or fewer out of 13 components accomplished the majority of weeks) = **6**	Less than 70% = **3**
Moderate	69-84% of goals met (i.e., 9 or fewer of 13 components accomplished the majority of weeks) = **6**	70%-83% = **11**
High	85% or higher of goals met (i.e., 11 to 13 out of the 13 components accomplished the majority of weeks) = **7**	over 84% = **5**

[23] Appendix C contains tables showing use of various *Saxon Phonics and Spelling* resources.

[24] It should be noted that although efforts were made to increase fidelity of implementation based on teacher log feedback, and as a whole moderate to high implementation was obtained, as a field study being conducted in real-world educational settings, there will always be some natural variation.

In addition, there was no evidence of contamination[25] in any of the sites. Treatment teachers would limit *Saxon Phonics and Spelling*-based conversations with teachers in their own group and used this program exclusively. When outside resources were used by control teachers, these never included the *Saxon Phonics and Spelling* program materials.

It should be noted that the potential for contamination was given careful consideration when determining the level of random assignment. Through years of research experience, PRES researchers have found that the benefits of random assignment at the teacher level (hence, controlling for school level factors) *with* careful monitoring of possible contamination, outweighs the risk of contamination. Procedures used to eliminate the threat of contamination included an in-depth study orientation with both treatment and control teachers, site visits made to both treatment and control classrooms to observe what was occurring in classrooms, and monthly teacher logs that monitored practices and materials used.

Overall, treatment teachers implemented the key Saxon Phonics and Spelling program components with low to high degrees of fidelity. The majority of teachers were able to complete over 70% of the program. In addition, there was no evidence of contamination between treatment and control groups. That is, the Saxon Phonics and Spelling program was delivered to treatment students only.

[25] Contamination refers to when the treatment (in this case, the *Saxon Phonics and Spelling* program) is also used in control classrooms and hence, the study becomes invalid.

RESULTS

This section is organized by the key evaluation questions and provides a summary of major findings first, followed by a more detailed account of the results. The findings described in the main body of this report provide a summary of overall conclusions that can be derived from the extensive analyses conducted. Detailed descriptions of the multiple analyses conducted on the assessment data are provided in Appendix E (pages 96-122).

Summary of Results

1. ***Do phonics, reading words, and spelling skills improve over the course of participating in the Saxon Phonics and Spelling program? Does this vary across different types of students and levels of implementation?***

Students using the *Saxon Phonics and Spelling* program significantly improved over the course of the school year in the areas of spelling, phonics, and reading words, including high frequency and sight words. Gains were also observed on developmental spelling stages. In particular, among 1st grade students who took the Morris & Perney Spelling test, results showed that students moved to higher spelling stages from pre- to post-testing.

Furthermore, the *Saxon Phonics and Spelling* program worked just as well with females and males, students who spoke a language other than English at home and those that did not, and special education and non-special education students. Among the remaining subgroups, differences in improvement between students were observed. Generally, Whites, 2nd graders, students not receiving free and reduced lunch, and lower-performing students

showed greater gains than minorities, 1st graders, students receiving free/reduced lunch, and higher-performing students, respectively. Nevertheless, among all subgroups, students using the *Saxon Phonics and Spelling* program showed significant gains in reading, phonics, and spelling.

Since there was some variation observed in overall implementation of the *Saxon Phonics and Spelling* program among treatment teachers, analyses were performed to examine if this affected student performance. Results showed that there was no significant relationship between overall *Saxon Phonics and Spelling* implementation levels and improved performance on the outcome measures. That is, students whose teachers used the *Saxon Phonics and Spelling* program improved on these measures, regardless of their level of overall implementation.

However, preliminary analyses of individual components of the *Saxon Phonics and Spelling* showed that there was a relationship between teacher's use of various *Saxon Phonics and Spelling* program components (e.g., use of worksheets and fluency readers) and gains in student phonics, reading, and spelling performance. Furthermore, the percent of lessons completed in the *Saxon Phonics and Spelling* program was a strong predictor of student gains in performance – the more lessons completed, the greater the improvement.

2. ***How does phonics, reading words, and spelling performance differ between students who use Saxon Phonics and Spelling as compared to students who do not use this program? Do effects on student achievement differ across types of students or settings?***

There were notable differences in treatment and control students' performance. Students using the *Saxon Phonics and Spelling* program showed more improvement than control students on the ITBS Word Analysis (which measures phonics and phonemic awareness skills), Spelling, and Reading Words tests. Most of the effect sizes, which provide an indication of the importance of results, would be considered educationally significant by the research literature.

In addition to treatment students outperforming control students on these measures, exploratory subgroup analyses showed a number of significant differences between treatment and control students who were females, Whites, African Americans, 2nd graders, receiving free/reduced lunch, attending various schools, and lower-performing. In particular, students in these subgroups that used *Saxon Phonics and Spelling* showed greater growth in performance from pre- to post-testing as compared to students that did not use this supplemental program.

3. ***Does participation in Saxon Phonics and Spelling result in other positive student outcomes (e.g., positive attitudes towards reading and so forth)?***

While the main focus of the *Saxon Phonics and Spelling* program is to improve upon important reading and spelling skills, other measures were included to explore if *Saxon Phonics and Spelling* was associated with positive impacts on student attitudes towards reading, phonics, and so forth. Results showed that, in general, treatment and control students had similar positive attitudes in regards to phonics, spelling, writing, motivation to do well in school, and perceived reading ability. However, control students had significantly more positive

attitudes towards reading as compared to treatment students during the Spring.

In terms of the program's effects on teachers, results showed notable increases in treatment teachers' levels of preparation and knowledge to teach the five elements of reading, spelling and writing, and their engagement in effective literacy practices from Fall to Spring. In addition, while during the Fall control teachers indicated having more knowledge, preparation, and engagement in effective literacy practices, treatment teachers caught up to control teachers in the areas of preparation and engagement in best practices by the Spring. Thus, there is evidence that suggests that the *Saxon Phonics and Spelling* program has a positive impact on teacher's level of preparation to teach phonics and spelling, and this in turn can lead to improvement in their pedagogical practices.

4. What did users of the Saxon Phonics and Spelling program think about it?

The *Saxon Phonics and Spelling* program was also highly regarded by the vast majority of teachers. A full 94% of treatment teachers surveyed agreed that that the program contributed to improved reading ability, and helped their students obtain greater phonics and spelling skills. A majority of teachers felt that the program provided them with the instructional background necessary to teach phonics and overall met their needs for both spelling and phonics instruction. In general, they also felt that the program was helpful in monitoring student progress. Generally, treatment students also liked the program. More than 70% indicated that they enjoyed the board work, and that the Wall Cards and letter/sound cards were helpful to them in learning and remembering phonics rules. The lowest rated item among students was the decodable readers.

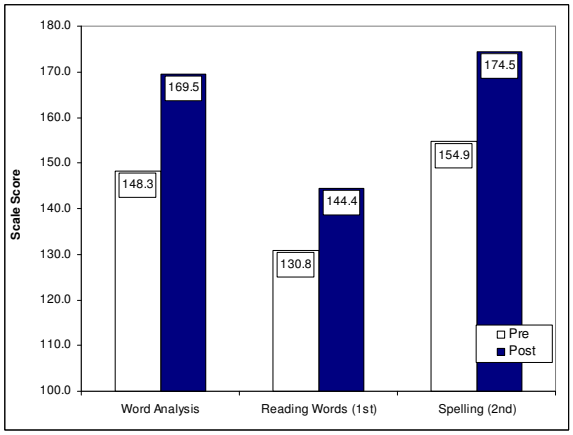
When asked what they felt were the greatest strengths of the *Saxon Phonics and Spelling* program, teachers' most often cited the Wall Cards, Review Decks and daily worksheets. Teachers also noted the sight word practice, the explicit phonics instruction, and the incremental approach to teaching phonics rules (i.e., building on student's knowledge throughout the school year) as being very beneficial to student performance. In terms of other programmatic feedback provided by teachers, they noted the following: (1) pacing of the lessons was at times unrealistic given the amount of material they had to cover; (2) the amount of time spent on coding words was sometimes overwhelming to students, caused confusion for students, and at times high-level students lost interest in the lesson; (3) a few noted that it took awhile for students (and themselves) to learn the program and become accustomed to the terminology, concepts, and activities employed by the program; and (4) some teachers also noted that the program did not integrate well with their basal reading and language arts program.

Detailed Results

1. DO PHONICS, READING WORDS, AND SPELLING SKILLS IMPROVE OVER THE COURSE OF PARTICIPATING IN THE SAXON PHONICS AND SPELLING PROGRAM? DOES THIS VARY ACROSS DIFFERENT TYPES OF STUDENTS AND LEVELS OF IMPLEMENTATION?

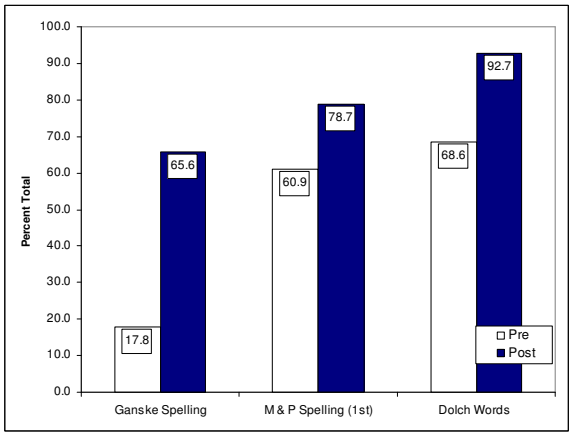
Multilevel modeling was performed on the data in order to measure growth from pre to post[26]. Results showed significant growth in treatment students as measured by all assessment measures, see Figures 2-3. That is, students who received the *Saxon Phonics and Spelling* program exhibited significant growth from pre- to post-testing on phonics, t-ratio(ITBS-Word analysis)=15.97, $p<.001$, reading words, t-ratio(ITBS-Reading words)=9.92, $p<.001$, t-ratio (Dolch Words)=14.16, $p<.001$, and spelling, t-ratio(ITBS-Spelling)=17.26, $p<.001$, t-ratio(Ganske spelling)=40.79, $p<.001$, t-ratio (M&P Spelling)=8.36, $p<.001$.

Figure 2. *Saxon Phonics and Spelling* Students' ITBS Performance at Pre and Post-testing



- There was significant growth on all ITBS subtests.

Figure 3. *Saxon Phonics and Spelling* Students' Performance on Dolch Word List, Ganske Spelling, and Morris & Perney (M&P) Spelling at Pre and Post-testing



- *Saxon Phonics and Spelling* students showed significant growth on the developmental spelling assessments as well as high frequency word reading via the Dolch test.

Improvement among *Saxon Phonics and Spelling* students can also be seen in growth of percentile ranks[27]. It is a general rule of thumb that if a student makes a year's growth for a year of instruction, then the percentile rank will remain the same. As shown in Table 6, the percentile rank on the ITBS assessment grew more than what would be expected in a typical academic year. Specifically, at post-testing, *Saxon Phonics and Spelling* students' percentile rankings increased, with the greatest jump occurring on the ITBS spelling test administered to 2nd grade students.

Table 6. *Saxon Phonics and Spelling* Students' Percentile Ranking (PR) at Pre and Post-testing

	Pre	Post
ITBS Word Analysis	47th	58th
ITBS Reading Words (1st)	50th	54th
ITBS Spelling (2nd)	46th	62nd

- In general, treatment students had higher percentile rankings at post-testing than pre-testing. Note that it is a general rule of thumb

[26] This analytical technique has the advantage of being able to take into account nesting (students within classrooms within teachers) and associated dependency issues (students within the same class being more similar to one another than to students in other classes). Detailed rationale on what statistical analyses was performed and why is provided on pages 107-122 in Appendix E, along with detailed statistics.

[27] Percentile ranks indicate the relative standing of a student in comparison with other students in the same grade in the norm (reference) groups (in this case, the nation) who took the test at a comparable time.

that if a student makes a year's growth for a year of instruction, then the percentile rank will remain the same. Therefore, students using *Saxon Phonics and Spelling* grew more than would be expected over the course of an academic year.

Treatment students' phonics, reading, and spelling performance improved significantly over the course of participating in the Saxon Phonics and Spelling program.

Gains were also observed as measured by developmental spelling stages. Among 1st grade students who took the Morris & Perney Spelling test, results showed that students moved to higher stages (see percents in green areas in Table 7). For example, 61.5% of students classified as phonetic during pretesting were at the transitional/correct stage at post-testing. In contrast, movement to lower stages was much less prevalent (see percents in red areas).

Table 7. *Saxon Phonics and Spelling* Students' Developmental Spelling Stage at Pre and Post-testing

		POST			
		Pre-phonetic	Semi-phonetic	Phonetic	Transitional/ Correct
PRE	Pre-phonetic	11.1%	11.1%	16.7%	61.1%
	Semi-phonetic	15.0%	15.0%	35.0%	35.0%
	Phonetic	.0%	9.2%	29.2%	61.5%
	Transition/ Correct	.0%	.0%	4.3%	95.7%

- Improvement among 1st *grade Saxon Phonics and Spelling* students was evident on the Morris and Perney Spelling Test. In particular, movement to higher spelling stages was more prevalent than movement to lower stages. For example, 61.5% of students classified as phonetic during pretesting were at the transitional/correct stage at post-testing.

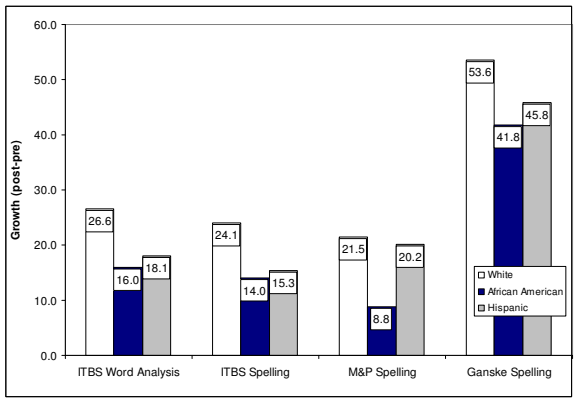
In order to examine whether the *Saxon Phonics and Spelling* program was associated with improvements among students of various subgroups, exploratory, descriptive analyses were conducted. Only the performance of *treatment* students in specific student populations (i.e. students with English as a second language, special education students, students receiving free/reduced lunch, females, ethnic minorities, and students at different grade levels and of various reading levels) was examined in these analyses. It should be noted that the sample sizes in the subgroups may be small and there are unequal sample sizes between those in the subpopulations and those not[28]. Therefore, with the caveat that these analyses are weak, this provides readers with preliminary, descriptive information on whether the program is associated with improvements among various subgroups.

Results showed that the amount of growth from pre- to post-testing was similar among females and males, students who spoke a language other than English at home and those that did not, and special education and non-special education students. That is, the program was associated with similar rates of improvement with all these subgroups.

However, differential growth rates were observed for the remaining subgroups. In terms of ethnicity, results showed that generally, Whites showed the greatest improvement, followed by Hispanics and African Americans as measured by the ITBS Word Analysis, ITBS Spelling, Morris & Perney Spelling, and Ganske Spelling tests, see Figure 4. However, all ethnic groups showed significant growth from pre- to post-testing, $p<.05$.

[28] The reader is referred to pages 108-115 in Appendix E for detailed statistics.

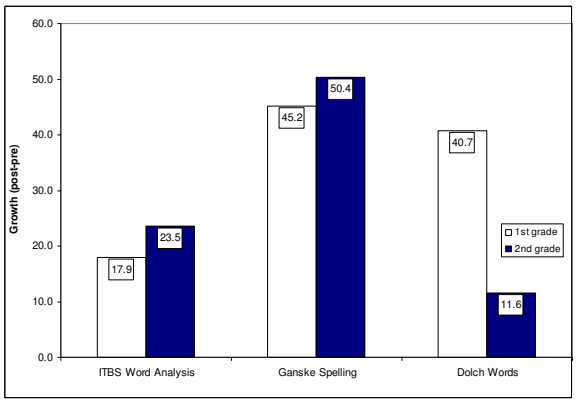
Figure 4. ***Saxon Phonics and Spelling*** **Students' Assessment Performance Gains by Ethnicity**



- While significant improvement was observed among students of various ethnic backgrounds, Whites tended to show the greatest improvement followed by Hispanics and then African Americans.

Analysis by grade level focused on those assessments that both 1st and 2nd graders took, namely the ITBS Word Analysis, Ganske Spelling, and Dolch Word tests. Results showed that 2nd grade *Saxon Phonics & Spelling* students showed greater change on the ITBS Word Analysis and Ganske Spelling tests as compared to 1st graders, see Figure 5. In contrast, 1st graders showed greater change than 2nd graders on the Dolch Word List test. However, this latter finding may be due to the high performance observed among 2nd graders at pretesting (Dolch words were noted as too easy) and therefore, there was less room for growth. On each of the assessments, both 1st and 2nd graders showed significant growth from pre- to posttesting on all measures, $p<.05$.

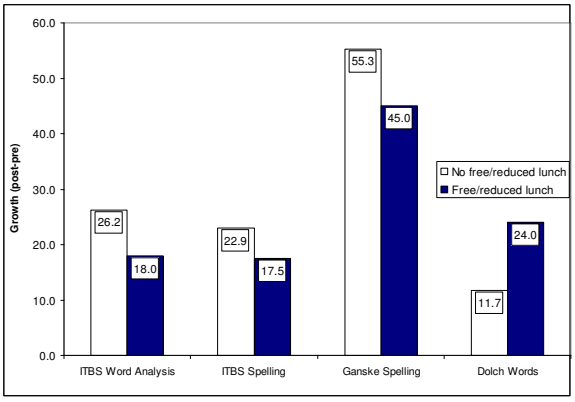
Figure 5. ***Saxon Phonics and Spelling*** **Students' Assessment Performance Gains by Grade Level**



- Second grade students demonstrated greater gains on the word analysis and Ganske Spelling test as compared to first grade students. In contrast, first grade students showed greater gains on the Dolch Word test than second graders. However, this may be due to the finding that second grade students performed very well at pretesting on this test and therefore, there was less room for growth as compared to first graders.

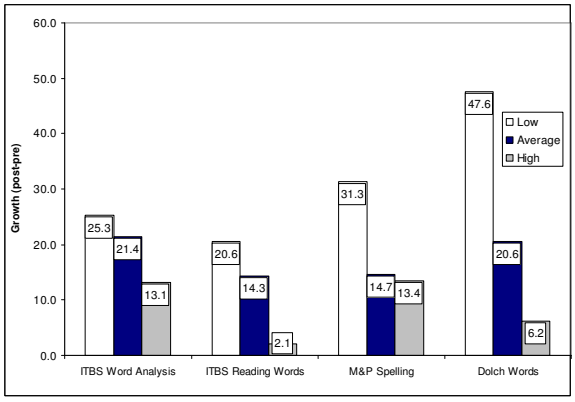
Analysis of free/reduced lunch status showed that *Saxon Phonics and Spelling* students not receiving free/reduced lunch showed greater change as measured by the ITBS Word Analysis, ITBS Spelling, and Ganske Spelling tests as compared to students receiving free/reduced lunch, see Figure 6. However, the reverse relationship was observed for the Dolch Word test. That is, students receiving free/reduced lunch showed greater change on high frequency word reading than students not receiving free/reduced lunch. It should also be noted that students in both subgroups showed significant growth from pre- to posttesting on all measures, $p<.05$.

Figure 6. *Saxon Phonics and Spelling* Students' Assessment Performance Gains by Free/Reduced Lunch Status



- Students receiving free/reduced lunch and those not receiving this assistance showed significant gains on all outcome measures. However, differential growth rates were also observed. Specifically, students not receiving free/reduced lunch showed greater gains on the ITBS Word Analysis, ITBS Spelling and Ganske Spelling tests. In contrast, students receiving free/reduced lunch showed greater gains on the Dolch Word test than students not receiving this aid.

Figure 7. *Saxon Phonics and Spelling* Students' Assessment Performance Gains by Beginning Literacy Levels



- While students at low, average and high literacy levels showed improvement in performance, lower-performing students tended to show the greatest amount of growth, followed by average and then high-performing students. It is important to note, however, that the observed relationship may be due to the fact that there is more room for improvement among the lower-level students as compared to higher performing students.

In order to categorize students on initial literacy levels, the ITBS percentile rankings at pretest were used. The bottom 30% were categorized as low, 31-69% were categorized as moderate, and the top 70% were categorized as high[29]. Comparisons were made between the three identified literacy levels. Results showed that the change in performance from pre to post was greatest among low-level students, followed by average and then high-level students as measured by the ITBS Word Analysis, ITBS Reading Words, Morris & Perney Spelling, and Dolch Word tests, see Figure 7. Note that this may be due to the greater room for improvement available for low-level students as compared to higher level students. In addition, all groups generally showed growth from pre- to post-testing on the assessments, $_{p<.05}$.

Overall, the Saxon Phonics and Spelling program worked just as well with females and males, students who spoke a language other than English at home and those that did not, and special education and non-special education students. However, there were also differences in improvement between students in other subgroups. Generally, Whites, 2nd graders, students not receiving free and reduced lunch, and lower-performing students showed greater gains than minorities, 1st graders, students receiving free/reduced lunch, and higher-performing students, respectively. Nevertheless, among all subgroups, students using the Saxon Phonics and Spelling program showed significant gains in reading, phonics, and spelling.

[29] This measure was used because it was the only standardized, norm-referenced test that could provide information on performance levels relative to a national sample.

In addition to these analyses among subgroups of *Saxon Phonics and Spelling* students, exploratory multilevel analyses on the relationship between overall levels of *Saxon Phonics and Spelling* implementation of components and student posttest performance were conducted. These analyses provide preliminary information on whether low, moderate, and high implementation fidelity of *Saxon Phonics and Spelling*[30] components was associated with student performance.

Results showed that there was no significant relationship between overall *Saxon Phonics and Spelling* implementation levels and improved performance on the outcome measures[31], $_{p>.05}$. That is, students whose teachers used the *Saxon Phonics and Spelling* program improved on these measures, regardless of their overall level of implementation.

Preliminary analyses showed that overall levels of Saxon Phonics and Spelling implementation was not related to improved performance in reading words, spelling, and phonics.

In order to obtain preliminary information on whether specific practices (including *Saxon Phonics and Spelling* components) were associated with positive student performance, the relationship between student's growth on the ITBS test and implementation of various *Saxon Phonics and Spelling* components and other teacher characteristics was examined. This enables researchers to examine the specific components that may be related to differing levels of change in student performance.

Analysis[32] showed that higher levels of change were associated with the following items:

- Percent of *Saxon Phonics & Spelling* lessons completed, $\beta=.38,\ p<0.001$
- Assigning the *Saxon Phonics & Spelling* worksheet to students, $\beta=.23,\ p<0.001$
- Engaging in effective reading and language arts practices, $\beta=.30,\ p<0.001$
- Using the *Saxon Phonics* spelling dictionary and reference booklets, $\beta=.20,\ p=.002$
- Using the *Saxon Phonics & Spelling* fluency readers, $\beta=.14,\ p=.026$

Note that with the exception of the extent to which teachers engaged in effective literacy practices, all items that significantly predicted gains in student performance were *Saxon Phonics and Spelling* program components. Indeed, the greatest predictor of improvement among the variables included in the analysis was the percent of *Saxon Phonics and Spelling* lessons completed. In contrast, time spent on reading, phonics, and spelling outside of the program did not significantly predict gains in performance.

It is important to note that strategies employed and program use, along with classroom characteristics, function *together* to help produce positive outcomes. Indeed, prior research has shown that pedagogical approaches employed by teachers are just as important as the intervention itself (Gersten, Lloyd & Baker, 1998). Hence, it is unsurprising that teacher engagement in effective pedagogical practices would be associated with student performance, along with program components.

[30] See section on Fidelity of Implementation for how this categorization was determined.
[31] Detailed statistics are presented on pages 115-116 in Appendix E.
[32] Detailed statistics are presented on pages 116 in Appendix E.

There was a relationship between teacher's use of various Saxon Phonics and Spelling program components (e.g., use of worksheets and fluency readers) and gains in student phonics, reading, and spelling performance. Furthermore, the percent of lessons completed in the Saxon Phonics and Spelling program was a strong predictor of student gains in performance.

Overall, these results on implementation suggest that although low to high levels of implementation of *Saxon Phonics and Spelling* components is associated with improved student reading performance, it is recommended that teachers aim to use this program on a daily basis. This is because increased use of the program itself, including some critical components as previously described, is associated with higher test scores.

Note that the focus of the above analyses was to examine if the *Saxon Phonics and Spelling* program was positively associated with student performance (and it is). These analyses do not examine how students improved over time as compared to control students. The following section presents analyses of how the treatment condition compares to the control condition.

2. DOES PHONICS, READING WORDS, AND SPELLING PERFORMANCE DIFFER BETWEEN STUDENTS WHO USE SAXON PHONICS AND SPELLING AS COMPARED TO STUDENTS WHO DO NOT USE THIS PROGRAM? DO EFFECTS ON STUDENT ACHIEVEMENT DIFFER ACROSS TYPES OF STUDENTS OR SETTINGS?

Prior to discussing the results found, it is important to reiterate that there were a number of similarities between control and treatment classrooms. Both types of classrooms taught similar content such as reading, phonics, spelling, writing, and so forth throughout the school year. In general, teaching styles and content taught were comparable across all treatment and control teachers. This is to be expected since both treatment and control teachers within sites were following the same standards and using the similar, if not the same, core reading and language arts curricula. The only notable difference between these classrooms throughout the year was the 50-60 minute explicit phonics instruction treatment group students received via *Saxon Phonics and Spelling*.

RESULTS

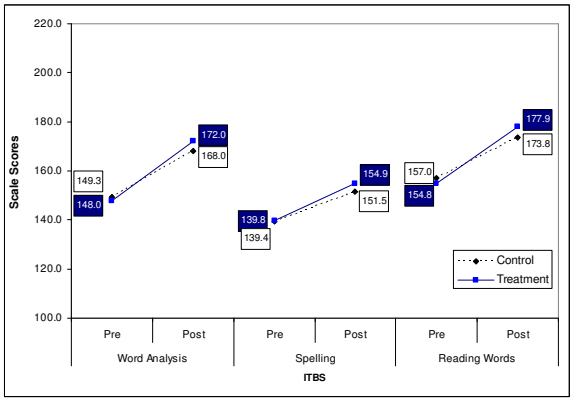
Multilevel models[33] were run to examine whether there was a significant difference in growth in performance between treatment and control students, as well as to account for statistical issues that can affect the validity of the results and to equate the groups on important variables (i.e., pretest, ethnicity, gender, grade, baseline teacher knowledge and preparation, baseline engagement in effective literacy practices, and school). Results of these analyses showed significant[34] differences between treatment and control students' performance as measured by the ITBS Word Analysis, $_{t\text{-ratio}=2.17,\ p=.03,\ d=.16}$, ITBS Spelling test, $_{t\text{-ratio}=2.28,\ p=.02,\ d=.30}$, and a marginally significant difference on the ITBS Reading Words subtests, $_{t\text{-ratio}=1.79,\ p=.07,\ d=.28}$. Specifically, on the ITBS Word Analysis, Spelling, and Reading Words subtests, treatment students showed greater improvement in these skills than control students, see Figure 8. However, on

[33] Detailed information and statistics regarding these results are presented on pages 117-119 in Appendix E.

[34] Significance levels have been adjusted for multiple comparisons using the procedure outlined by Sankoh and colleagues (1997). Significant differences are those with $p<.03$ and marginally significant differences are those with $p<.07$. Appendix F (pg. 126) discusses this adjustment.

the Dolch Word List, Ganske Spelling, and Morris & Perney Spelling tests, no significant differences were observed between treatment and control students, $p>.05$.

Figure 8. Pre and Post Performance on ITBS Subtests by Group



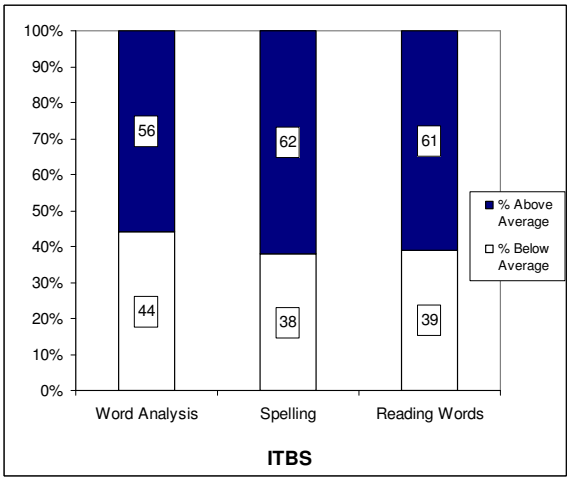
Results showed that treatment students performed significantly better than control students on the ITBS Word Analysis and Spelling tests, and marginally better on the Reading Words test.

Effect size[35] is a commonly used measure of the importance of the effect of an intervention (in this case, the *Saxon Phonics and Spelling* program). Given the similarity in treatment and control classrooms, prior research[36], and the fact that the duration of the study and exposure to the program occurred during *one* school year, small (.20) to moderate (.50) effect sizes were expected. It should also be noted that according to Slavin (1986), a leader in educational research, an effect size of .25 is considered educationally significant (i.e. it has a real, meaningful impact on achievement).

The effect sizes obtained were 0.16 for word analysis, 0.30 for spelling, and 0.28 for reading words. The latter two measures exceed the 0.25 threshold for educational significance put forth in the research literature.

In order to better understand the effect sizes found[37], Figure 9 displays the percent of treatment students that can be expected to be *above* the average of the control group (see blue part of bar). As shown, 56%, 62%, and 61% of students using the *Saxon Phonics and Spelling* program are more likely to have scored above the average of control students as measured by the ITBS Word Analysis, Spelling, and Reading Words scale scores respectively.

Figure 9. Percent of *Saxon Phonics and Spelling* Students Above and Below Average Relative to Control Students (ITBS)



- Results show that 56%, 62%, and 61% of treatment students scored above the average control student on the ITBS Word Analysis, Spelling, and Reading Words scale scores respectively.

[35] More specifically, effect size (ES) provides a measure of the relative position of one group to another. For example, with a moderate effect size of d=.5, we expect that about 69% of cases in Group 2 are above the mean of Group 1, whereas for a small effect of d=.2 this figure would be 58% and for a large effect of d=.8 this would be 79%.

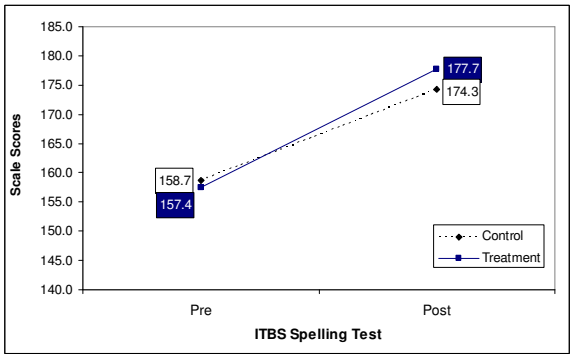
[36] Our prior experience indicates that supplemental programs are likely associated with small effect sizes (.19 to .35).

[37] This information is displayed in addition to modeled estimates presented in Figure 8 because multi-level models produce coefficients based on complex algorithms that are not easily interpreted, especially with the inclusion of multiple covariates.

To examine if there were differences in performance between different subgroups of treatment and control students, subgroup effects were analyzed via multilevel modeling for subgroups that had a sample size of 61 or greater[38] (i.e., gender, grade, ethnic status, free/reduced lunch status, and school). As previously noted, multilevel models account for statistical issues that can affect the validity of the results. Furthermore, it is important to view this analysis as exploratory[39].

Results showed a significant program effect for females on the ITBS Spelling test, t-ratio=2.11, p=.04. This means that females exposed to the *Saxon Phonics and Spelling* program showed significantly higher growth on ITBS Spelling than females not exposed to this program, see Figure 10. No other gender differences on assessment measures were observed.

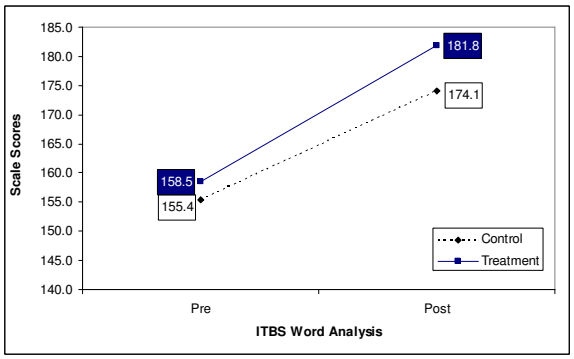
Figure 10. Females Pre and Post Performance on the ITBS Spelling Test by Group



- Female students who used the *Saxon Phonics and Spelling* program showed significantly higher growth on the ITBS Spelling test than females not exposed to the program.

Significantly greater growth on the ITBS Word Analysis test was also observed among 2nd graders who used the *Saxon Phonics and Spelling* program as compared to 2nd graders not using the program, t-ratio=2.07, p=.04, see Figure 11. No other differences by grade level were observed.

Figure 11. Second Graders Pre and Post Performance on the ITBS Word Analysis Test by Group



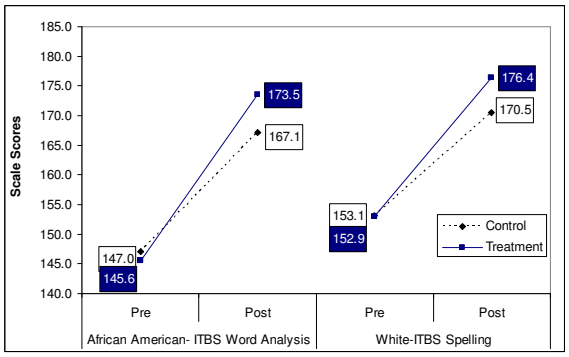
- Second grade students who used the *Saxon Phonics and Spelling* program showed significantly higher growth on the ITBS Word Analysis (phonics) test than 2nd graders not exposed to the program.

For African Americans and Whites, results showed significant improvement among these students who used *Saxon Phonics and Spelling* program as compared to African Americans and Whites not using the program, as measured by the ITBS Word Analysis, t-ratio=1.90, p=.06, and ITBS Spelling test, respectively, t-ratio=2.44, p=.02, see Figure 12.

[38] This is in accordance with the procedures employed by NAEP researchers (Swinton et al., 2001). This number was obtained by determining the sample size necessary to detect an effect size of .5 and have a power of .8.

[39] Detailed information on why this is exploratory and non-casual and statistics regarding these results are presented on pages 120-121 in Appendix E. Marginally significant results are included above.

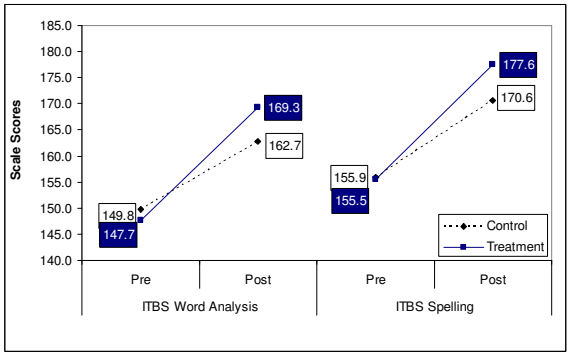
Figure 12. African Americans and Whites' Pre and Post Performance on the ITBS Word Analysis and Spelling Tests by Group



- African Americans and Whites who used the *Saxon Phonics and Spelling* program showed significantly more improvement on the ITBS Word Analysis and Spelling test, respectively, than African Americans and Whites who did not use the program.

Students receiving free/reduced lunch also were positively impacted by the *Saxon Phonics and Spelling* program. In particular, these students showed higher growth in performance on the ITBS Word Analysis, t-ratio=2.91, p=.004, and ITBS Spelling tests, t-ratio=2.26, p=.02, than students that did not use this program, see Figure 13.

Figure 13. Free/Reduced Lunch Students' Pre and Post Performance on the ITBS Word Analysis and Spelling Tests by Group



- Students receiving free/reduced lunch and who used the *Saxon Phonics and Spelling* program showed greater improvement on the ITBS Word Analysis and Spelling tests as compared to these students who did not use the program.

A number of positive effects were also observed at the school level. Specifically, students at school D that used *Saxon Phonics and Spelling* showed significantly better performance on the ITBS Word Analysis, t-ratio=2.95, p=.004, ITBS Spelling, t-ratio=2.92, p=.004, Ganske Spelling, t-ratio=4.69, p<.001, and Dolch Word List, t-ratio=2.01, p=.04, than students who did not receive this supplemental instruction, see Figures 14-15. Possible explanations for the number of significant differences observed at this school include: (1) implementation was very high for one of the treatment teachers, and (2) the school had a number of lower-performing and ESL students participating in the study, and anecdotal and statistical analysis reveals that this program may be more effective for students in these populations.

Figure 14. School D Students' Pre and Post Performance on the ITBS Word Analysis and Spelling Tests by Group

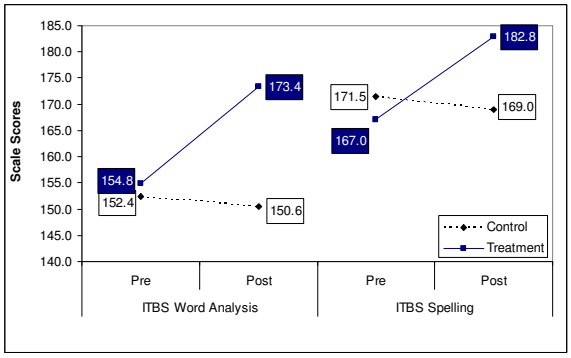
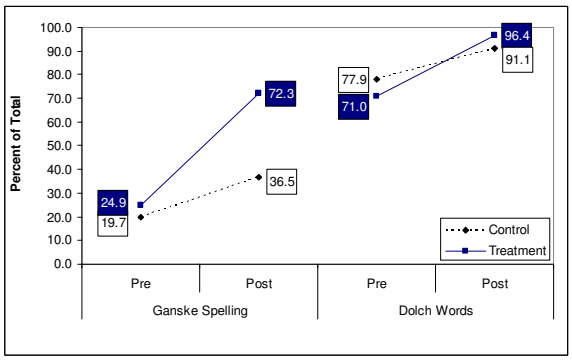


Figure 15. School D Students' Pre and Post Performance on the Ganske Spelling and Dolch Word Tests by Group



- Students at school D who used *Saxon Phonics and Spelling* showed significantly better performance on the ITBS Word Analysis, ITBS Spelling, Ganske Spelling, and Dolch Word tests than students who did not receive this supplemental instruction.

There were also positive differences observed at other schools. *Saxon Phonics and Spelling s*tudents at school F showed significantly higher growth on the ITBS Word Analysis, t-ratio=2.86, p=.005, ITBS Reading Words, t-ratio=2.19, p=.03, and the Ganske Spelling tests, t-ratio=2.82, p=.005, than students who did not receive this supplemental instruction, see Figures 16-17. It should be noted that treatment teachers at this school were characterized as having high implementation levels. In addition, *Saxon Phonics and Spelling* students at schools B and E also showed greater growth than students at these schools who were not exposed to this program, as measured by the Ganske Spelling test, t-ratio=1.92, p=.05 and t-ratio=2.21, p=.03, respectively, see Figure 18.

Figure 16. School F Students' Pre and Post Performance on the ITBS Word Analysis and Reading Words Tests by Group

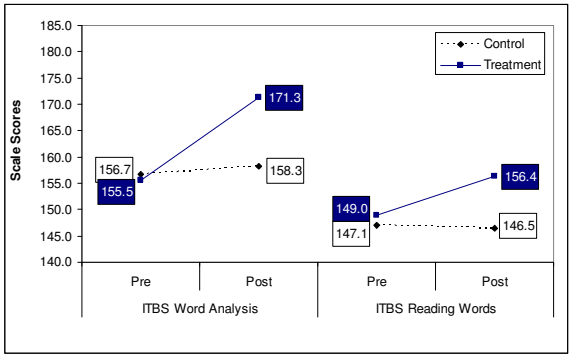
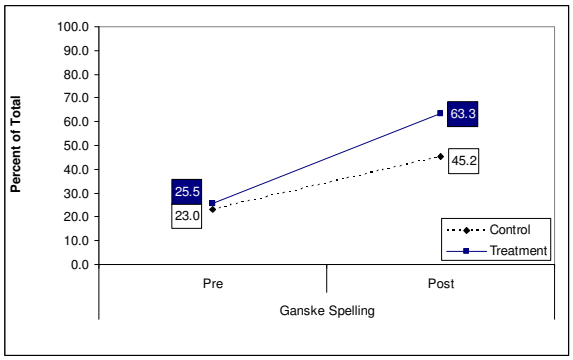
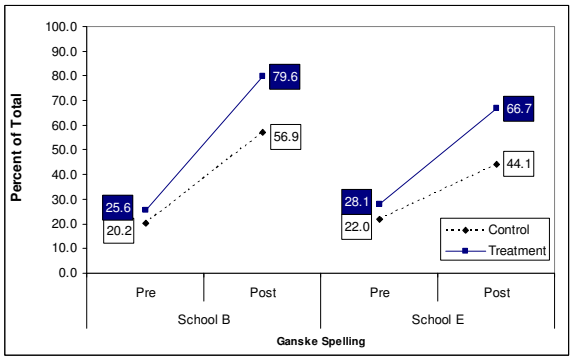


Figure 17. School F Students' Pre and Post Performance on the Ganske Spelling Test by Group



- Students at school F who used *Saxon Phonics and Spelling* showed significantly greater growth on the ITBS Word Analysis, ITBS Reading Words, and Ganske Spelling tests than students who did not receive this supplemental instruction.

Figure 18. School E and B Students' Pre and Post Performance on the Ganske Spelling Tests by Group



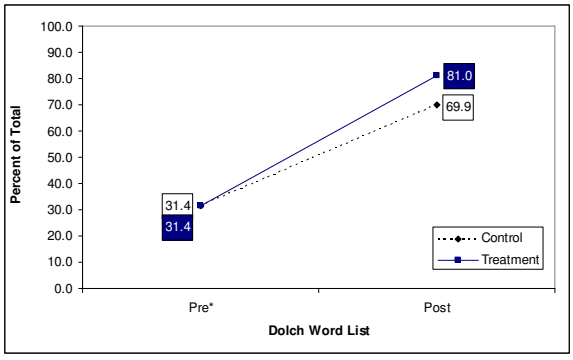
- Students at schools E and F who used *Saxon Phonics and Spelling* showed significantly greater growth on the Ganske Spelling tests than students who did not use *Saxon Phonics and Spelling*.

Further exploratory analyses (ANCOVA) were performed on the subgroups with sample sizes less than 61 (i.e., English language status, special education, and literacy levels). It is important to reiterate that due to the small sample sizes and other statistical issues, no causal, conclusive statements should be made. Nevertheless, these results are

presented for preliminary, exploratory purposes.

Results showed no significant differences between treatment and control students who spoke a language other than English at home, were in special education, and who were at high or average literacy levels *after* controlling for pretest differences, as measured by all assessment measures. This means that there was no difference between treatment and control students in phonics and spelling skill levels at post-testing among these subgroups of students. However, among low literacy level students, there was a significant difference as measured by the Dolch Word test. Specifically, low performing treatment students had higher scores than low performing control students after equating the students on pretest scores, $F(1, 114)=4.84$, $p=.03$, see Figure 19.

Figure 19. Low Literacy Level Students' Pre and Post Performance on the Dolch Word List by Group



*Covariate pretest value.

- Among low literacy students, students using *Saxon Phonics and Spelling* program had significantly higher growth on the Dolch Word test as compared to students not using this supplemental program.

Results showed a number of significant differences between treatment and control students who were females, Whites, African Americans, 2nd graders, receiving free/reduced lunch, attending various schools, and lower-performing. In particular, students in these subgroups that used Saxon Phonics and Spelling showed greater growth in performance from pre- to post-testing as compared to students that did not use this supplemental program.

3. How Does Participation in Saxon Phonics and Spelling Result in Other Positive Student Outcomes (e.g., Positive Attitudes Towards Reading and So Forth)?

In order to explore if *Saxon Phonics and Spelling* was associated with other positive student and teacher outcomes, pre- and post-surveys were administered to measure changes in attitudes towards reading, spelling, phonics-related activities and teacher practices.

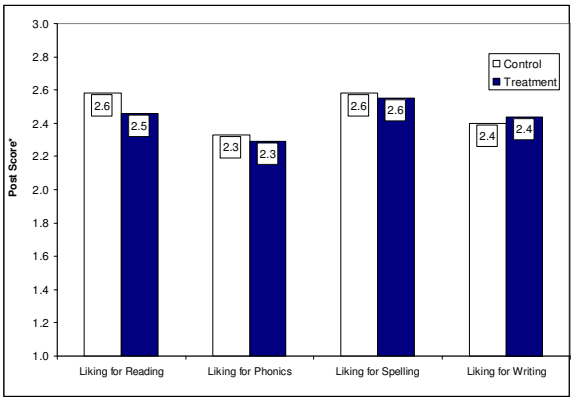
Student Attitudes

Comparisons[40] between liking for reading, spelling, writing, and phonics activities of control and treatment students showed that, for the most part, both groups displayed similar, positive attitudes towards spelling, writing, and phonics, see Figure 20. However, a significant difference was observed, *after* controlling for pre-attitudes. Specifically, control students indicated liking reading more than treatment students. Recall that the attitude scales ranged from 1 (disagree) to 3 (agree).

[40] Statistics for these results are as follows: (1) positive attitudes towards reading, $F(1, 608)=7.61$, $p=.006$, $d=.13$, (2) phonics, $F(1, 608)=1.13$, $p=.29$, $d=.08$, (3) spelling, $F(1, 607)=.39$, $p=.53$, $d=.06$, (4) writing, , $F(1, 603)=.58$, $p=.45$, $d=.06$, (5) perceived levels of reading ability, $F(1, 609)=3.30$, $p=.07$, $d=.12$, and (6) academic effort/motivation, $F(1, 609)=3.42$, $p=.07$, $d=.13$.

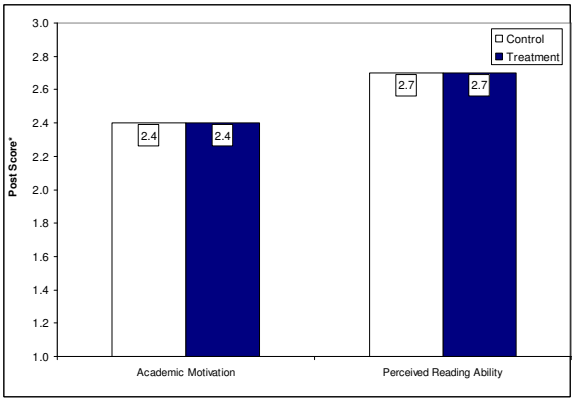
In addition, motivation in academics (e.g., wanting to do well in school) and perceived reading skills were similar across treatment and control students, see Figure 21.

Figure 20. Student Affective Attitudes by Group



*Adjusted for pre-attitude.

Figure 21. Student Attitudes about Reading Ability and Effort/Motivation by Group



*Adjusted for pre-attitude.

In general, treatment and control students had similar positive attitudes in regards to phonics, spelling, writing, motivation to do well in school, and perceived reading ability. However, control students had significantly more positive attitudes towards reading.

Teacher Practices, Knowledge, and Preparation for Five Elements of Reading, Spelling and Writing

As previously noted, treatment and control teachers differed significantly at baseline in terms of their preparation and knowledge of the five elements of reading, spelling, and writing, as well as engagement in effective literacy practices, with control teachers demonstrating more of these dimensions than treatment teachers. In order to examine if these differences diminished, analysis were conducted to examine whether or not there was more growth in these areas among treatment teachers as compared to control teachers. In other words, analysis focused on examining if changes occurred among treatment teachers so that they were able to "catch up" with control teachers in terms of their knowledge, preparation, and engagement in effective literacy practices.

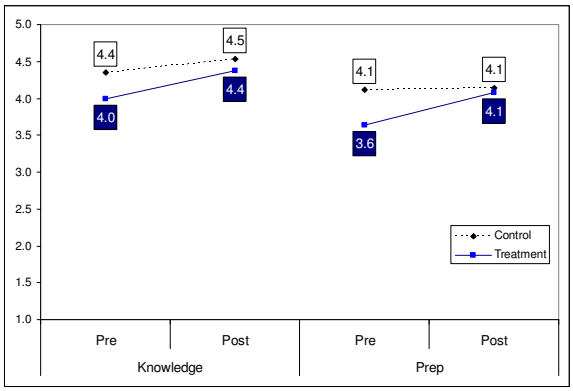
Results showed two significant effects. First, treatment teachers showed significantly greater improvement in their preparation to teach the five elements of reading (i.e., phonics, phonemic awareness, fluency, comprehension, and fluency), spelling and writing as compared to control teachers, $F(1, 26)=4.88$, $p=.04$, $d=.69$, see Figure 22[41]. Second, treatment teachers showed significantly greater improvement in their engagement in effective literacy practices than control teachers $F(1, 26)=8.17$, $p=.008$, $d=.85$, see Figure 23. Indeed, during the Spring, there were no longer significant differences between control and treatment teachers in terms of preparation or engagement in effective literacy practices. These findings are notable because even with the small sample size (n=35) and associated low power, these findings were significant and effect sizes were large (d=.69 to .85). As

[41] These items are based on scale from 1-5 with 1 being not at all prepared and 5 begin strong preparation.

such, this indicates that this is a major positive finding with regard to the supportive resources provided by the program in terms of helping teachers improve upon their pedagogical practices.

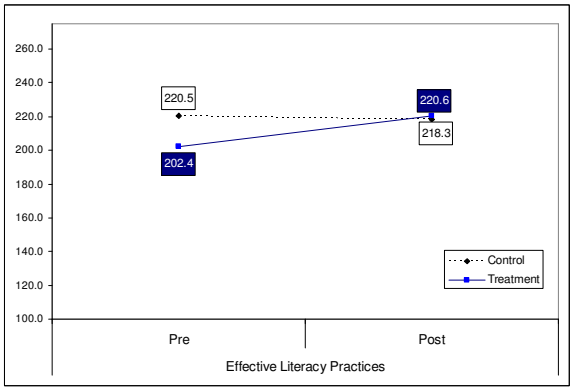
In addition, although greater improvement was observed among treatment teachers in terms of knowledge, the level of change was not significantly different to the change observed among control teachers, $p>.05$.

Figure 22. Teacher Knowledge and Preparation to Teach Elements of Reading, Spelling, and Writing by Group



- Treatment teachers showed significantly greater improvement in their level of preparation to teach the five elements of reading, spelling, and writing as compared to control teachers. While control teachers indicated greater preparation in the Fall, treatment teachers "caught up" in terms of their preparation in the Spring.

Figure 23. Teacher Engagement in Effective Literacy Practices by Group



- Treatment teachers showed significantly greater improvement in their engagement in effective literacy practices as compared to control teachers. While control teachers indicated greater engagement in the Fall, treatment teachers "caught up" by the Spring.

> *"I used to teach 3rd grade so I didn't really know how to explicitly teach phonics…I learned all the rules and names of [phonics components] that I didn't know before. You have a lot of teachers who need the explicit phonics instruction as well."*
>
> *-- 2nd Grade Teacher*

There were increases in treatment teachers' levels of preparation and knowledge of the five elements of reading, spelling and writing and their engagement in effective literacy practices from Fall to Spring. In addition, while during the Fall, control teachers indicated having more knowledge, preparation, and engagement in effective literacy practices, treatment teachers caught up to control teachers in the areas of preparation and engagement in best practices in the Spring.

4. WHAT DID USERS OF THE SAXON PHONICS AND SPELLING PROGRAM THINK OF THE PROGRAM?

> *"I love it [Saxon Phonics and Spelling program] - in six years of teaching, this is the first time I feel like I can teach reading. I'm feeling comfortable teaching reading."*
>
> *-- 2nd grade teacher*

Treatment teachers overall, liked the program. Anecdotal information obtained from treatment teachers indicated that the

Saxon Phonics and Spelling program met their instructional needs in the areas of phonics and spelling, and ultimately helped to improve students reading skills.

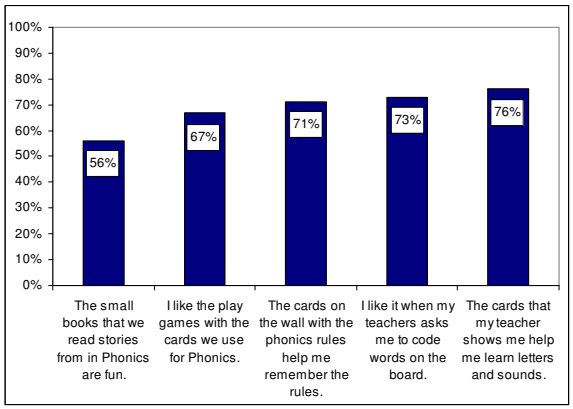
> *"They [students] are getting things they wouldn't with just the basal."*
> *-- 2nd Grade Teacher*

> *"Spelling has really improved with the help of sight words…amazing! I'm pleased overall and the growth is noticeable.*
> *-- 2nd Grade Teacher*

A vast majority of treatment teachers thought the program met their instructional needs and helped their students gain better reading, spelling and phonics skills.

Generally, treatment students also liked the program, see Figure 24. As shown, more than 70% indicated that they enjoyed board work, and that the Wall Cards and letter/sound cards were helpful to them in learning and remembering phonics rules. Approximately 67% of students noted that they enjoyed playing games with the Kid Cards. The lowest rated item (56%) referred to the stories in the decodable readers. However, it is important to note that many treatment teachers noted that they had limited time to allow students to play Kid Card games or to use the decodable readers. As such, this may in part reflect the more limited time that students were exposed to these activities.

Figure 24. Student Attitudes about the *Saxon Phonics and Spelling* Program



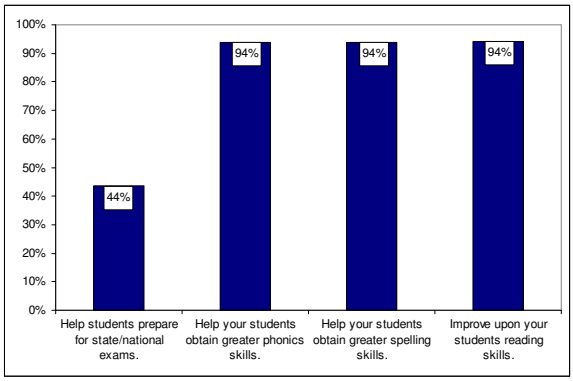
> *"I like the* board work*, and I think the students do - in the younger ages they like the chance to go up to the overhead and the chalkboard. I think the board work is definitely beneficial."*
> *--2nd grade teacher*

As shown in Figure 25, teachers overwhelmingly felt that the program had a positive effect on students. Indeed, 94% agreed that that the program contributed to improved reading ability, and helped their students obtain greater phonics and spelling skills. Only 44% of teachers, however, agreed that the *Saxon Phonics and Spelling* program helped them to prepare their students for state and/or national testing.

> *"The children have learned a lot of coding and how to attack words by using these strategies."*
> *--1st grade teacher*

> *"It helps them understand the [phonics] rules of why a letter sounds a certain way. For example, rather than just memorize that /ō/ is long, it tells them WHY it's long, why it makes the long sound."*
> *--1st grade teacher*

Figure 25. Teacher Attitudes about the *Saxon Phonics and Spelling* Program's Effect on Students (% Indicating Agreement)



When asked what they felt were the greatest strengths of the *Saxon Phonics and Spelling* program, teachers' most often cited the Wall Cards, Review Decks and daily worksheets. Teachers also noted the sight word practice, the explicit phonics instruction, and the incremental approach to teaching phonics rules (i.e., building on student's knowledge throughout the school year) as being very beneficial to students.

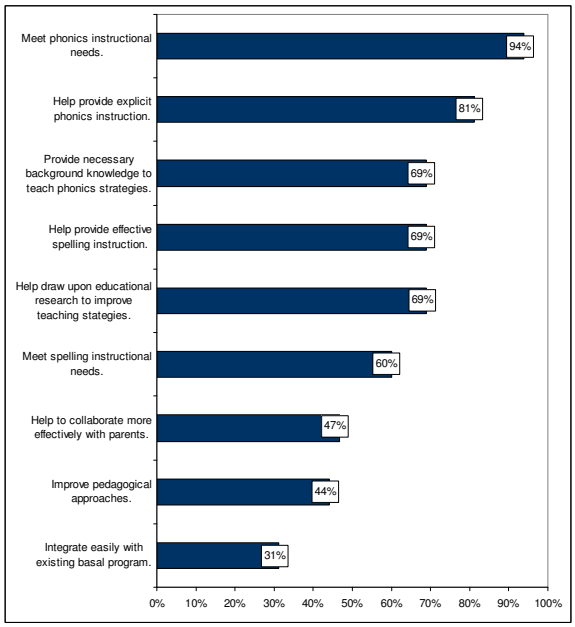
> *"I liked the cards, definitely. The sight words as well as the picture cards particularly, because they do very well visually…those parts of the program - it just gets them more excited because it makes it more of a game out of it."*
>
> *--2nd grade teacher*

A majority of teachers felt that the program provided them with the instructional background necessary to teach phonics and overall met their needs for both spelling and phonics instruction, see Figure 26. Note that while fewer teachers (60%) said that the program met their instructional needs in spelling, not all treatment teachers were using the spelling component of the *Saxon Phonics and Spelling* program with their students during the study. The lowest rated item, with 31% agreement, pertained to whether the program integrated easily with the basal program being used.

> *"It is sometimes hard to integrate the program with the basal reading and spelling we already have in place as part of our core language arts program."*
>
> *--1st grade teacher*

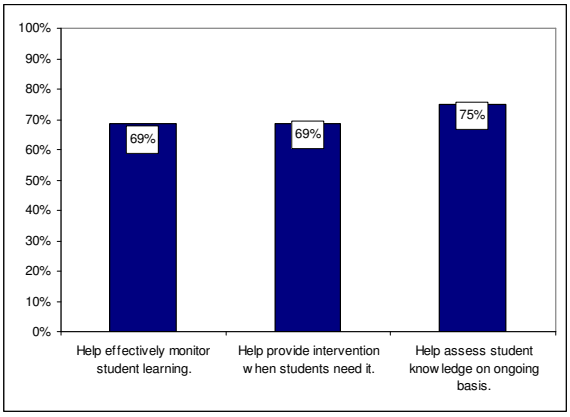
> *"I just think that it would have amazing results if we could get a basal (language arts program) that went along with the Saxon Phonics!"*
>
> *--1st grade teacher*

Figure 26. Teacher Attitudes of the *Saxon Phonics and Spelling* Program (% Indicating Agreement)



Teachers, in general, felt that the program was helpful in monitoring student progress. Among teachers using the *Saxon Phonics and Spelling* program, 75% indicated that the program was helpful in assessing student knowledge on an ongoing basis, see Figure 27. In addition, 69% felt that the program helped to monitor student learning and provide intervention when students needed it.

Figure 27. Teacher Attitudes of Progress Monitoring Mechanisims in the *Saxon Phonics and Spelling* Program (% Indicating Agreement)



Teacher ratings of various program components are provided in Appendix C. The program as a whole was rated as useful by more than 87% of teachers. The components that teachers perceived as most useful included:

- Review Decks
- Teacher's Manual
- Teacher's Resource Binder
- Wall Cards/Posters
- Student Decodable Readers
- The *Saxon* Phonics daily worksheet
- Assessments provided with program.

> *"I like the way each lesson is bundled together. It tells you exactly what they need to do so it is real user-friendly (for subs and student teachers as well)."*
>
> *--1st grade teacher*

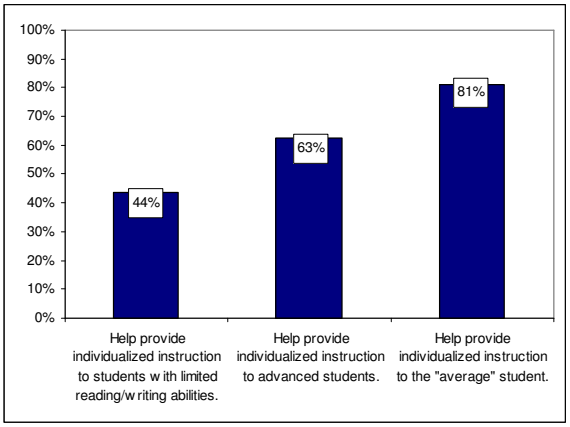
Teachers liked the overall layout of the Teacher's Manual and the accessibility of having each lesson as its own booklet and easily accessible. Teachers felt that the program was well organized and that the sequence of the lessons provided a nice flow from one lesson to the next.

Generally, teachers agreed that the program was useful in providing instruction to the "average student" (81%), see Figure 28. While 63% of teachers felt the program helped to provide individualized instruction to advanced students, only 44% thought it was helpful in providing individualized instruction to students with limited reading/writing abilities. However, during interviews, several teachers noted that they felt that the program was very helpful for lower-performing students.

> *"I think it's really helpful for students with low confidence in reading. Now they TRY words they don't know - now they know they can figure it out!"*
>
> *--1st grade teachers*

> *"I really like the Saxon program itself. It works very well - it certainly helps the struggling readers become much more independent and it really helps them along…because it really helps them break down the words, it teaches them how to recognize the chunks in the words rather than try to sound out every letter, and become more familiar with the diagraphs and the different sounds that they see."*
>
> *--2nd grade teacher*

Figure 28. Teacher Perceptions of Assistance Provided by *Saxon Phonics and Spelling* for Individualized Instruction (% Indicating Agreement)



There were two primary areas of weakness identified by teachers in the *Saxon Phonics and Spelling* program. First, teachers noted that the pacing of the lessons was sometimes unrealistic. They felt that the program had a lot of material in each lesson which was difficult to cover in the allotted time, especially during the latter part of the year when there was more review and skills taught were more sophisticated. Time constraints in general were reported to have been one of the major barriers that teachers faced when implementing the program. Particularly, teachers noted pacing issues with the classroom practice portion of the lesson (using the Kid Cards) and also using the decodable and fluency readers.

> *"There is not enough time in the school day to use everything Saxon offers or suggests."*
>
> *--1st grade teacher*

The second weakness teachers identified was the amount of coding required in the program. Teachers felt that the amount of time spent on coding words was sometimes overwhelming to students, caused confusion for students, and at times students lost interest in the lesson when there was a lot of coding involved. Specifically, teachers noted high ability students losing interest during coding activities while lower-performing students would get frustrated. It was suggested that for higher ability students, who sometimes lose interest in coding and other parts of the lesson, that the next grade level of the product be used to continue to engage their interest.

> *"I think too much time is spent on coding the words and not enough time actually looking at words that have the sounds being introduced."*
>
> *--2nd grade teacher*

In terms of other programmatic feedback provided by teachers, a few noted that it took awhile for students (and themselves) to learn the program. That is, it took time for them to become accustomed to the terminology, concepts, and activities employed by the program. As previously noted, some teachers also noted that the program did not integrate well with their basal reading and language arts program. Still, teachers made attempts to integrate the skills and concepts taught in *Saxon Phonics and Spelling* into their reading and language arts time, as well as other academic areas. As one teacher noted:

> *"I use the Saxon strategies in the guided reading or basal program. For example, anytime they come across a word they do not know in any subject, I use different word attack strategies to help them read."*
>
> *-- 1st grade teacher*

Overall, Saxon Phonics and Spelling teachers and students liked the program. Specifically, the majority of teachers cited that the program met their instructional needs for phonics. They also noted that the incremental approach and daily worksheet were very important in helping students' succeed in reading, phonics, and spelling.

CONCLUSION

The results obtained from this randomized control trial indicate that the *Saxon Phonics and Spelling* program is significantly related to positive student outcomes. Students using the program showed significant growth in phonics, spelling, and reading words from pre- to post-testing. Moreover, significant differences were observed between treatment and control students' performance. Students using the *Saxon Phonics and Spelling* program showed more improvement than control students on the ITBS Word Analysis, (which measures phonics), Spelling, and Reading Words tests. Results also showed a number of significant differences between treatment and control students who were females, Whites, African Americans, 2nd graders, receiving free/reduced lunch, attending various schools, and lower-performing. In particular, students in these subgroups that used *Saxon Phonics and Spelling* showed greater growth in performance from pre- to post-testing as compared to students that did not use this supplemental program.

Given the similarity between treatment and control classrooms' core reading/ language arts curricula and that, on average, only 24 more minutes of daily instruction in phonics and spelling was noted, this lends confidence that differences observed between treatment and control students are due to the presence and absence of the supplemental *Saxon Phonics and Spelling* program and not the core reading/language arts curricula.

There were also notable increases in treatment teachers' levels of preparation and knowledge of the five elements of reading, spelling and writing, and their engagement in effective literacy practices from Fall to Spring. In addition, while during the Fall control teachers indicated having more knowledge, preparation, and engagement in effective literacy practices, treatment teachers caught up to control teachers in the areas of preparation and engagement in best practices by the Spring. Thus, there is evidence that suggests that the *Saxon Phonics and Spelling* program has a positive impact on teacher's level of preparation to teach phonics and spelling, and this in turn can lead to improvement in their pedagogical practices. Such findings are noteworthy because while phonics instruction is important, oftentimes, these are areas where teacher training is lacking.

The *Saxon Phonics and Spelling* program was also highly regarded by a number of teachers. A full 94% of treatment teachers surveyed agreed that that the program contributed to improved reading ability, and helped their students obtain greater phonics and spelling skills. A majority of teachers felt that the program provided them with the instructional background necessary to teach phonics and overall met their needs for both spelling and phonics instruction. In general, they also felt that the program was helpful in monitoring student progress. Generally, treatment students also liked the program. More than

70% indicated that they enjoyed board work, and that the Wall Cards and letter/sound cards were helpful to them in learning and remembering phonics rules. The lowest rated item among students (56%) referred to the stories in the decodable readers.

The observed effects were associated with small effects (.16-.30). However, the effect sizes obtained for reading words and spelling can be considered educationally significant *despite* the following limitations:

a) This was a new program for the schools and they only implemented it for one school year. In fact, three schools[42] began implementation a full two months after the start of the school year. This meant that these schools had less exposure to the *Saxon Phonics and Spelling* program as compared to those who were able to implement earlier towards the start of the school year. Furthermore, program effects take time to develop as teachers become more familiar with the program and its resources. Lastly, this program was only implemented for approximately 4-5 hours per week.

a) With the exception of school B, the reading/language arts content that was taught in all of these classes was very similar. The most notable difference between treatment and control classrooms was the addition of the *Saxon Phonics and Spelling* program. Similarities in the core reading/language arts programs or supplemental phonics programs and *Saxon Phonics and Spelling* can diminish program effects that may be associated with the *Saxon* program.

b) There were some treatment teachers who did not implement all program components with fidelity. The components that teachers had more difficulty implementing included the language/alphabet activity during warm-up, the classroom practice activity, use of the student spelling dictionary and reference booklets during new increment, weekly sight word evaluation, and completing 50 to 60 minutes of daily *Saxon Phonics and Spelling* instruction.

In conclusion, this RCT with its use of quantitative and qualitative methods enabled PRES Associates to determine that the *Saxon Phonics and Spelling* program produced more positive outcomes relative to classrooms that did not supplement their core reading program and it caused positive improvements in student performance in the areas of spelling, phonics, and reading words. Teachers and students also enjoyed using the program. In addition, researchers were able to obtain information on how the program was used in real-world classrooms, as well as preliminary information on factors that contributed to the effects observed. Still, further research is needed to build upon the findings presented in this report.

[42] Schools C, D, and E were late additions to the study.

REFERENCES

Chall, J., & Popp, H. (1996). *Teaching and Assessing phonics: Why, What, When, How*. Cambridge, MA: Educators Publishing Service.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin.

Ganske, Kathy Ann (1994) Developmental spelling analysis: A diagnostic measure for instruction and research. Ph.D. dissertation, University of Virginia, United States -- Virginia. Retrieved June 2, 2007, from ProQuest Digital Dissertations database. (Publication No. AAT 9425756).

Gersten, R., Lloyd, J. W., & Baker, S. (1998). *Designing high quality research in special education: Group experimental designs*. Washington DC: U.S. Department of Education.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.

Morris, Darrell & Perney, Jan (1984). Developmental spelling as a predictor of first-grade reading achievement. *The Elementary School Journal*, 84(4), 440-457.

National Reading Panel. (April, 2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Retrieved online September 7, 2006 from http://www.nichd.nih.gov/publications/nrp/smallbook.htm.

Raudenbush, S. & Bryk, A. (2002). *Hierarchical linear models for social and behavioural research: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

Raudenbush, S., Spybrook, J., Liu, X.-f., & Congdon, R. (2005). *Optimal design for longitudinal and multilevel research: Documentation for the Optimal Design Software*. New York: William T. Grant Foundation.

Sankoh, A. J, Huque, M. F., & Dubey, S.D. (1997). *SISA Bonferroni.* Southampton: D.G. Uitenbroek. Retrieved online July 20, 2006 from http://home.clara.net/sisa/bonhlp.htm.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analysis and traditional reviews. *Educational Researcher*, *15*, 5-11.

Slavin, R. E., Karweit, N. L, & Madden, N. A. (1989). *Effective programs for students at risk.* Allyn & Bacon: Boston.

Swinton, S. S., Freund, D. S., & Allen, N. L. (2001). Conventions used in hypothesis testing and reporting NAEP results. In Allen, N. L., Donoghue, J. R. & Schoeps, T. L. (Eds.), *The NAEP 1998 Technical Report* (pp. 247-254). Retrieved online June 12, 2005 from National Center of Education Statistics Web site: http://nces.ed.gov/nationsreportcard/pdf/main1998/2001509b.pdf.

Appendix A

Crosswalk between Study Design Characteristics and WWC Review Standards

Table A1: Crosswalk between *Saxon Phonics and Spelling* Study Design Characteristics and WWC Review Standards

WWC Causal Evidence Standards	Study Characteristics	Reference
Randomization: Were participants placed into groups randomly?	Teachers were randomly assigned to control and treatment groups within schools[43]. Random assignment was conducted via random numbers tables by PRES researcher.	Final Report, pg. 10
Baseline Equivalence: Were the groups comparable at baseline, or was incomparability addressed by the study authors and reflected in the effect size estimate?	There were no baseline differences according to pretest assessments. In terms of student background variables, results showed significantly more 2nd graders and less 1st graders in the *Saxon* group as compared to the control condition. In addition, there were a higher proportion of control students classified as being high performing in terms of literacy as compared to *Saxon* students. However, significant differences were observed in terms of teacher knowledge and preparation to teach via the five elements of reading, and spelling and writing, and their engagement in effective literacy practices. At baseline, control teachers reported having more knowledge and preparation for teaching important elements of reading, spelling, and writing, and engaged in effective literacy practices to a greater extent than treatment teachers. As a result, care was taken to include the variables that were non-equivalent between the treatment and control groups as covariates in the multilevel models. These are reflected in effect size estimates.	Technical Appendix E, pgs. 100-101
Differential Attrition: Is there a differential attrition problem that is not accounted for in the analysis?	Both measurement and dropout attrition was examined. ▪ The lack of Dolch post-testing by one teacher resulted in less control students taking the Dolch Word test as compared to *Saxon* Phonics students. However, there were no *performance* differences between group and those who completed tests and those that did not. ▪ While dropout attrition was associated with lower scores on three tests, this was consistent across both groups (differential attrition was not observed).	Technical Appendix E, pgs. 103-106

[43] There are a number of reasons why random assignment to treatment conditions was done at the teacher level within schools. The most important reason for selecting this level of assignment is that such a design helps to establish causality by reducing the threat that school-level factors could have potentially contributed to differences between treatment and control groups. That is, school "A" might have had something else going on (besides the treatment) that may have influenced student performance on the outcome measures. Since treatment and control groups were within the same school, school-level explanations of differences were reduced. Another reason for within school assignment is that it is likely that the treatment and control groups will possess similar characteristics at the onset of the study and therefore enhance comparability. Third, one of the criteria put forth by the DIAD study is that treatment and control groups need to be drawn from the same local pool (Valentine & Cooper, 2003). The definition of local pool provided in this study refers to subjects within the same classroom or school. According to the criteria, randomization at the district level would not be drawing people from the same local pool. Note, while this may increase the potential threat of contamination this was contained by an in-depth study orientation, monthly teacher logs, and site visits (see pages 15-18). Notably while random assignment at the teacher level within schools helps researchers control for school level differences as potential explanations of observed differences between treatment and control groups, teacher level factors can also be present and are important predictors of student performance (Gersten, Lloyd, & Baker, 1998). Though random assignment at the teacher level should help address this, with smaller sample sizes it is less likely that group equivalence will be ensured. In order to address this potential threat to initial group equivalence, additional data was collected on teacher background and classroom practices and examined and taken into account in interpretation of results. The reason why random assignment was *not* done at the lower levels (i.e., within classrooms or at the student level), was because (1) the threat of contamination if the same teacher taught both curricula was considered too great, and (2) it is not practical to randomly assign students to conditions. Indeed, schools rarely allow outside researchers to randomly assign students to use one program over another and use of the school's scheduling system as a source for randomization is not acceptable due to the fact that students' previous class schedule, ability level, student/parental requests, and so forth are factored in; this is not random.

WWC Causal Evidence Standards	Study Characteristics	Reference
Overall Attrition: Is there a severe overall attrition problem that is not accounted for in the analysis?	Overall attrition due to student dropouts was 9%. Note that this was part of the initial site selection criteria; in order to minimize attrition, historical mobility rates were examined and sites with high attrition rates were eliminated from consideration.	Technical Appendix E, pgs. 104-106
Disruption: Is there evidence of a changed expectancy/ novelty/disruption, a local history event, or any other intervention contaminants?	There was no evidence of changed disruption, or a local history event. Contamination among control group teachers was also not observed. Potential treatment contaminants included: 1) the less than desirable implementation of some components of the program, such as classroom practice and the spelling component, and 2) the late addition of three sites (began implementation approximately 2 months after the start of school year. Together, these likely had a negative impact on size of treatment effects observed.	Final Report, pg. 31 and 52
Intervention Fidelity: *1. Documentation*: Is the intervention described at a level of detail that would allow its replication by other implementers? *2. Fidelity*: Is there evidence that the intervention was implemented in a manner similar to the way it was defined?	**1. Documentation:** The implementation guidelines provided in Appendix D clearly outline the expectations for implementation of the program. The *Saxon Phonics and Spelling* program is described herein in sufficient detail and references for further documentation from the publisher are provided. **2. Fidelity:** Extensive procedures were put in place to measure fidelity of intervention including training, implementation guidelines, monthly teacher logs, and site visits. While there was evidence for low implementation of *Saxon Phonics and Spelling* components, overall fidelity of implementation can be characterized as moderate.	1. Final report pgs. 18-19, 27 and Appendix D 2. Final report pg. 30
Outcome Measures: 1. *Reliability:* Is there evidence that the scores on the outcome measure were acceptably reliable? 2. *Alignment:* Is there evidence that the outcome measure was over aligned to the intervention?	**1. Reliability:** The assessments used are reliable and valid. ITBS developers report split-half reliability estimates between .80-.89. Ganske report's test-retest reliability at .94. The Morris & Perney spelling test is correlated with the MAT8 (.61). The reader is referred to the developers of these assessments and ITBS technical manual for more detailed information on psychometric properties. **2. Alignment**: These tests measure phonics, spelling and ability to read and identify words based on cues. These are considered the key dimensions that the program focuses on.	1. Final report pgs. 12-14 2. Final report pgs. 12-14

WWC Causal Evidence Standards	Study Characteristics	Reference
People, Settings, and Timing: 1. *Outcome Timing*: Does the study measure the outcome at a time appropriate for capturing the intervention's effect? 2. *Subgroup Variation*: Does the study include important variations in subgroups? 3. *Setting Variation:* Does the study include important variations in study settings? 4. *Outcome Variation:* Does the study include important variations in study outcomes?	**1. Outcome Timing:** Post measures were taken within 1 month of the end of the school year. At three sites, pretest measures were taken within 1 month of the beginning of the school year. At the remaining three sites, due to their late addition into the study, pretesting occurred approximately 2 months after the beginning of the school year. **2. Subgroup Variation:** The sample includes variations in gender, race/ethnicity, free/reduced lunch status, and grade level. Analyses were conducted by all subgroups, although small sample sizes English language status, special education, and literacy level means that the results by subpopulations should be interpreted with caution. **3. Setting Variation:** One site was in a large urban city (Georgia). Four other sites were suburban and in Texas, Oklahoma and Idaho. Another site was in a rural Indiana community. All schools were public with an enrollment (377-680 students) that is typical of schools at this level and in such settings (see page 22 for site characteristics table). **4. Outcome Variation**: Subtests measuring phonics, spelling, and reading were used to measure the effect of the program on student performance. In addition, exploratory analyses on the relationship between the program and student and teacher attitudes, and classroom practices was also examined.	1. Final Report pg. 12 2. Final Report pg. 23 3. Final Report pgs. 21-22. 4. Final Report pgs. 12-15
Testing Within Subgroups: 1. *Analysis by Subgroup*: Can effects be estimated for important subgroups of participants? 2. *Analysis by Setting*: Can effects be estimated for important variations in settings? 3. *Analysis by Outcome Measures*: Can effects be estimated for important variations in outcomes? 4. *Analysis by Type of Implementation*: Can effects be estimated for important variations in the intervention?	**1. Analysis by Subgroup:** Effects were estimated via multilevel models for the subpopulations that we had sufficient data for (i.e., gender, grade level, ethnicity, free/reduced lunch status). In addition, analysis using special education students, students whose primary language was other than English, and students at various reading levels was limited to a small sample. **2. Analysis by Setting:** Analyses by setting consisted of examining program effects by school. These analyses showed a significant interaction with schools. In particular, the program was found to be effective in schools D and F, given the number of significant program effects observed. **3. Analysis by Outcome Measures:** Effects were estimated for each subtest as well as for affective outcomes. **4. Analysis by Type of Implementation:** Program effects were estimated by variations in implementation. Results showed no relationship between implementation levels and student performance.	1. Technical Appendix E pgs. 120-121 2. Technical Appendix E pgs. 120-121 3. Technical Appendix E pgs. 118-119 4. Technical Appendix E pgs. 115-116

WWC Causal Evidence Standards	Study Characteristics	Reference
Analysis: 1. *Statistical Independence*: Are the students statistically independent or, if there is dependence, can it be addressed in the analysis? 2. *Statistical Assumptions*: Are statistical assumptions necessary for analysis met? 3. *Precision of Estimate*: Is the sample large enough for sufficiently precise estimates of effects?	**1. Statistical Independence:** Analysis of the intraclass correlations showed that dependency was an issue among this sample of students. However, this was addressed by using hierarchical linear modeling and inclusion of covariates. **2. Statistical Assumptions:** All underlying statistical assumptions were met. **3. Precision of Estimate:** Power analyses revealed that multilevel models have enough power to detect medium to large effects.	1. Technical Appendix E pg. 99 2. Technical Appendix E footnote on pg. 99 3. Technical Appendix E pgs. 101-102
Reporting: 1. *Complete Reporting*: Are findings reported for most of the important measured outcomes? 2. *Formula*: Can effects be estimated using the standard formula (or an algebraic equivalent)?	**1. Complete Reporting**: All main findings for the outcomes are presented in the Technical Appendix E. **2. Formula:** All effect sizes (Cohen's d) for outcomes measures are calculated, using method described in Appendix F, and presented in the report.	1. Technical Appendix E 2. Technical Appendix E pg. 119

Appendix B

Case Study of Site Visits

Elementary School A

About the School – Elementary school A is located in a large city in Oklahoma, surrounded by a lower-middle class suburban neighborhood. The school itself is over 50 years old and in relatively poor condition, with exposed piping, poor lighting and inadequate ventilation. Currently servicing grades K-5, preparations are underway to include 6th grade in the near future. Enrollment during the 2006-2007 school year was 377, with a student-teacher ratio of 16.

Results from the 2006 Oklahoma Core Curriculum Test (OCCT) show that 68% of third graders were at or above satisfactory in reading, as compared to the statewide average of 82%. Additionally, the state Department of Education did not identify this school as in need of improvement for 2005-2006.

The student population is diverse:

- 50% Hispanic;
- 28% White;
- 11% African-American;
- 6% American-Indian;
- 3% Asian/Pacific Islander.

Approximately 83% of the students were eligible for free or reduced-priced lunches, and 41% were classified as English-Language Learners.

About the Participants – Four classes participated in the study: one treatment and one control class at both the first and second grade level. A total of four teachers and 91 students took part, with an average class size of 26 ranging from 22 to 29.

All teachers devoted 2 to 2.5 hours a day to reading and language arts instruction. Instruction occurred in the mornings, although teachers noted that they would conclude lessons in the afternoons if needed. *Saxon* lessons were conducted in the mornings amidst the language arts block, usually beginning around 9:00 am. Treatment teachers reported spending close to 90 minutes on *Saxon* Phonics lessons at the beginning of the year, decreasing this time to 60 minutes by the end of the year.

The second grade teachers described their classes as having a very broad range of abilities in terms of performance and average in comparison to other classes at their grade level. The first grade control teacher indicated that students in this class were higher performing than the other first grade class (treatment); however, both noted that their classes were average overall.

The highly transient, highly bilingual student population was noted as being challenging. Student engagement across the four participating classrooms was moderate, and student behavior issues often resulted in lost instructional time, with teachers halting instruction for classroom management reasons. This was true across all four participating classes.

About the Reading and Language Arts Curriculum and Resources – Teachers instructed reading and language arts with a basal series (2001). Several teachers commented that this basal program did not have enough explicit instruction or a strong phonics component, but that it was mandated by the school district that all teachers use the program as a primary instructional resource. Additionally, the district curriculum map drove instruction at the elementary level, and as such most teachers used the basal almost exclusively for planning purposes. The first grade treatment teacher, however, noted that the basal was not used as much as the *Saxon* decodable readers for reading instruction in that classroom, and that while the curriculum map was followed, the basal was used more as a supplement. Furthermore, the second grade control teacher noted that seasonal stories and lesson plans would take precedence over the basal when applicable or when the stories in the basal simply did not suit the class.

Second grade teachers noted that they usually met once each week to coordinate on upcoming language arts skills and reading and were on pace with one another. In contrast, the first grade teachers did not co-plan because of different instructional techniques and planning resources. Furthermore, the two treatment teachers would meet regularly to discuss overall use of the *Saxon* program and exchange ideas.

In addition to the basal, all teachers used supplemental materials to augment instructional resources and target specific student needs. Among them, a published phonics program was in place in both first grade classrooms, including the treatment class where it was used prior to *Saxon* for phonics at the beginning of the year and as a supplement after *Saxon* was introduced. This program was not used in the second grade treatment class as a supplement, nor was it used for phonics instruction in the second grade control classroom, as the teacher described it as too easy for second grade students. Instead, the second grade control teacher employed a skill-based workbook emphasizing phonics skills including long vowels, diphthongs, and contractions. Both treatment teachers also supplemented with another skill-based program (2002) targeting language arts skills such as suffixes, word families, vowel pairs and homophones. The first grade control teacher also used Read, Write, Sing and Spell, a technique that puts information to music to help students learn and was formerly a building-wide initiative, as well as a workbook-website tandem that provides practice in both reading and writing. All teachers at School A reported the use of various teacher-created materials as well.

The site liaison and reading specialist in the building would also pull students into the reading lab for intervention five times a week for approximately 25 minutes. Treatment and control students in first and second grade would receive additional instruction on phonics and sight words. The reading specialist used any number of resources available, including texts and blackline masters. The reading specialist would also integrate social studies and science concepts into reading.

Several school-wide reading initiatives were in place to support student reading efforts at School A. Namely, there was a reading challenge initiated by the principal in which students were encouraged to read books to earn tickets that could be redeemed for prizes. Local businesses helped participate in the Book Buddies program by offering rewards such as pizza and ice cream. Additionally, students who read 100 books or more attended an end of the year "bounce party," and were eligible for a drawing to win $100.

About the Reading and Language Arts Classrooms – All treatment and control classrooms were adequate in space and had sufficient resources. Computers were available in all four classrooms for students and teachers and were used frequently. Libraries for independent reading were available and well-organized within the classrooms.

A typical basal lesson at the first grade level began with whole group reading, as the teacher would read aloud for 20 minutes each day. During reading, the teacher would ask comprehension questions and activate prior knowledge while moving through the story. The students would then engage in an activity after the reading, usually either a phonics worksheet or a hands-on activity related to the content of the lesson. As students were prone to getting off-task easily, both teachers were very explicit with instructions, clarifying expectations, and asking questions to check for understanding. Once students were engaged, the teacher would walk among student tables and assist where necessary, typically with struggling students. Time permitting, the class would finish with whole group practice on spelling words where students would spell words chorally.

Overall, both second grade teachers engaged in similar practices. A typical basal lesson at the second grade level varied depending on the lesson between whole group instruction and rotating small groups. Whole group lessons would begin with students seated at their desks as the teacher introduced the story or lesson activity. Prior to reading, the teacher might review vocabulary words using the board or overhead or activate previously taught skills through group discussion. Then students would begin reading aloud one at a time while the teacher assisted with the pronunciation of difficult words. Afterwards, the teacher would introduce an activity for students to complete independently at their seats or in small groups related to the content of the lesson. Small group instruction as a lesson would begin with very explicit instructions by the teacher as to assignments and expectations, with constant references to the behavioral guidelines in the classroom. After students had their assignments, the teacher would allow student movement to stations, and would meet with an individual student or two for fluency observations. The remainder of the class was either following along with a story at the listening station, practicing language arts skills on the computer, completing a writing assignment at his or her desk, or reading independently in preparation for a fluency conference with the teacher.

As noted earlier, during classroom observations, students at School A were only moderately engaged in lessons, were often disruptive and needed constant redirection. Students engaged in work were equally needy, requesting individual attention on a regular basis and frequently left their seats to pursue the teacher. In the event the teacher's attention was compromised, students would begin off-task behavior, and the level of disruption in the classroom would continuously rise until the teacher addressed all of the students with either behavioral consequences or work incentives. Thus, teachers at both grade levels would often lose precious instructional time as students transitioned from one activity to the next or as they refocused students on work.

While there was no school-wide policy, homework was assigned at least 4 nights a week at both grade levels although the amount varied depending on the teacher. Assignments ranged anywhere from 10 to 30 minutes a night and consisted of reading to parents, writing sentences using weekly words, and studying for spelling or vocabulary assessments. Several participating

teachers noted that there was little home support, which effected students' efforts at home. Treatment teachers reported assigning *Saxon* homework infrequently, as the work was usually attended to in class.

About the *Saxon Phonics and Spelling* Classrooms – Both first and second grade treatment teachers were using the *Saxon* program daily and completing one lesson each day, barring irregular schedules. These teachers taught *Saxon Phonics and Spelling* within their morning reading and language arts instructional period, and would also use either the basal phonics component or a supplemental resource in addition to the *Saxon* program.

The first grade treatment teacher followed the Teacher's Manual mostly as outlined. Spending 60 minutes a day, a typical lesson involved student movement as a transition between activities within the lesson, to prevent students from getting "antsy," according to the teacher. While seated on the carpet towards the back of the classroom, students would review the Deck Cards with the teacher and complete the language/alphabet warm-up activity. Although the teacher commented that this was not always done towards the end of the year because the students "got it," it would take almost 20 minutes when transition time was included. Students would return to their seats to complete the continual review section of their worksheet, which took another 20 minutes, before moving back to the carpet for introduction to the new increment. At this point, students would actively participate in the board work before moving back to their seats one last time to work on the application section of the worksheet independently. At this point the teacher would visit students needing additional help or requiring attention to keep focus. Time permitting, the teacher would incorporate the fluency and decodable readers into the lesson as recommended. However, due to time constraints this teacher would omit the readers as well as the Kid Cards and portions of the Deck Cards. Further, this teacher did use the spelling portion of the *Saxon* program as it came up, but noted that the class was not on a regular weekly schedule and thus did not have a weekly spelling list or assessment.

The second grade treatment teacher was more exact in adhering to the Teacher's Guide regarding *Saxon* implementation. Although the class frequently skipped portions of the warm-up due to time, a typical lesson would begin with an alphabet activity taking approximately 10 minutes. The new increment was introduced thoroughly, as students would refer to posted rule and spelling charts around the room and reference student dictionaries as called for in the lesson. When applicable, students would engage in board work and always completed the application portion of the worksheet independently. At the end of the instructional period, the teacher would correct students' papers individually and assign the remaining sections as homework. While most components were used on a regular basis when recommended within the lesson, the Kid Cards, handwriting component and site word evaluations were skipped more often due to issues of time. The spelling component was used five days a week, and students would complete the review regularly.

Highlights – Although student engagement was moderate and behavior issues abounded, teachers at School A did a relatively good job keeping students interested in all lessons and were committed to providing meaningful instruction to all students. All four participating teachers were capable, and treatment teachers did a good job implementing the *Saxon* program.

Elementary School B

About the School – Elementary school B is located in a city in Southeastern Idaho. The community is suburban and the school is in a residential, middle-class area. The school itself is older, but well-kept. It serves grades K-6. Total enrollment for School B was 610 in the 2005-2006 school year, and the student-teacher ratio was 20.

During the 2005-2006 statewide testing, 91% of third graders were at or above proficient on the Idaho Standards Achievement Test as compared to the state average of 84%. This school made adequate yearly progress in the prior two school years.

The student population is primarily white:

- 85% White;
- 10% Hispanics
- 2% African American;
- <1% American Indian or Alaskan Native;
- 1% Asian/Pacific Islander

Approximately 42% of the students were eligible for free or reduced-priced lunches.

About the Participants – Participants in the study from elementary school B consisted of only 2nd grade classes. There were approximately 65 students and 3 teachers (2 treatment and 1 control) at the second grade level who participated in the study. The average class size was 22 students, with a range from 22 to 23.

For reading instruction, students were grouped by ability level. *Saxon* instruction occurred in the students' homeroom classes. While these homeroom classes were not intentionally ability grouped, teachers indicated that the groupings were fairly homogenous. Particularly, between the two treatment classes, one group was predominantly lower-ability and the other was a predominantly higher-ability class. The remaining control class was characterized as containing a more broad mix of students and was classified as "high-average" in terms of ability.

The participating study teachers indicated that they did not have a specified time to meet and discuss lesson planning or instructional strategies, but that this occurred informally once to twice per month. Based on observation and interviews at the site, it is apparent that teachers focused on their own classrooms and typically did not collaborate in terms of pacing or activities. All teachers followed a district guideline and taught to the same state content standards.

About the Reading and Language Arts Curriculum – As previously noted, students were divided by ability for reading instruction. There were four groups of students: low, low-average, high-average and high. The control teacher provided instruction to the high-average group and the treatment teachers to the low-average and high groups. The teacher with lowest group of reading students did not participate in the study. There were no school-wide reading and/or language arts initiatives taking place.

Each teacher used a variety of materials for their reading classes. The treatment class with high ability students relied primarily on an older (1993) basal program. This program is comprised of two anthologies with 3-4 units each. Materials with the program include six student anthologies and the teacher's guide; according to the teacher, other materials have dwindled over the years. The program focuses instruction on engaging students through authentic, meaningful literature. Before reading, the teacher previews, questions and encourages students to make predictions. During reading the focus is on connecting ideas, becoming personally involved and using cues to comprehend. After reading, the teacher guides and encourages students to respond and reflect.

The two other teachers (one treatment and one control) used a newer (2002) program as their basal curriculum. The treatment teacher with the low-average group used this as her main resource, while the control teacher indicated that she did not rely much on this program, instead incorporating many of her own resources after years of teaching practice. The 2002 curriculum provides built-in test preparation, leveled resources designed to reach all learners, and explicit instruction. The program is designed for each week to begin with a phonics story, followed by a main reading selection which targets the weekly comprehension skill and vocabulary, then ends the week with language arts tied to the reading selection, including a writing portion. There are also leveled readers with the program that were used at the site.

As indicated, the control teacher, while sometimes pulling resources from the basal, primarily used her own teacher-collected materials. These included a variety of worksheets, trade books, and class activities. In addition to the basal and her own materials, the control teacher also regularly used the Accelerated Reader (AR) program where students read a book independently and then take a computer test on the reading. Students were assigned to read one book and take the AR test each week. This was a big focus for the teacher and her students, and a very measurable achievement for the class.

All teachers taught reading and language arts for a one hour block, during the mid-morning, five days a week. Phonics instruction took place in the homeroom classes and the time of day varied. The control teacher noted that she placed moderate emphasis on phonics in her homeroom class and used a published phonics program as a guide, but did not follow it closely.

About the Reading and Language Arts Classrooms - All treatment and control classes were adequate in space and had sufficient materials for all students. All classrooms had some independent reading books available in a classroom library, however, these libraries were limited in variety and genre. Teachers said they regularly took students to the school's library to select independent reading materials.

Based on observations, teachers all began their lessons in the same format, with warm-up exercises such as a review of previous material or a quick discussion of a previous days assignment. The warm-up took approximately 10 minutes. The lesson was then taught to the class as a whole group for the next 30 minutes. Activities during this instruction time varied, but typically consisted of reading selections (either from the text or a teacher selected source) and engaging in activities related to the reading such as group discussion, writing responses to questions or reviewing vocabulary in the passage. Teachers noted they rarely taught using small

group instruction as their classes were already grouped homogenously. Teachers rarely engaged in enrichment activities for advanced students or support for lower-level students, due to the nature of whole group instruction. Throughout the lesson, almost all of the students were fully engaged, especially when cued by the teacher, and classroom management did not seem to be a hindrance at any point. Independent practice was observed, although teachers noted during interviews that this was not always typical. All teachers appeared to be equally effective in implementing the basal lessons. Each teacher also had 2-4 parent volunteers who came into their classes once each week to help with reading. They would listen to students read individually, or partner read.

While the above denotes a typical language arts/reading lesson as observed between the three teachers, a typical lesson for the control teacher is difficult to summarize as she tended to use a variety of activities and did not consistently follow a pattern in her instruction. She noted that for most lessons she tried to include a short phonics lesson everyday, and then sometimes she would have students do "tracking," reading along to a book on tape, or read independently for the AR test. She also taught and reviewed spelling each day with her homeroom students, using a variety of teaching styles to encourage the use of phonetic sounds throughout her lessons.

Both control and treatment teachers report that they assigned independent reading exercises in-class 4-5 times per week. Students were required to do 20 minutes of independent reading as homework. Additionally, treatment teachers said they assign the *Saxon* homework worksheet regularly, although it is sometimes completed in class rather than taken home. Teachers (both treatment and control) reported that 95-100% of all homework is turned in. One treatment teacher said that she checked the homework, but did not grade it, while the other teacher regularly graded the homework assignments. Teachers indicated that they used a variety of ways to evaluate students including frequent informal assessment such as observation and oral questioning. Other assessment modes included in-class quizzes on spelling and vocabulary, the computer generated AR tests (for the control class), and a biannual district assessment for all classes.

About the *Saxon Phonics and Spelling* Classrooms – Treatment teachers at this school worked to fully implement each aspect of the program and generally followed each step as outlined in the implementation guidelines. Both treatment teachers began the class by reviewing the deck cards. However during the spring observation one teacher reviewed the cards at the end of the lesson rather than the beginning. It was observed that the teachers effectively introduced the new lesson increment and discussed the new card with their classes. Additionally both teachers engaged their students in the board work section of the lessons. Following the board work, teachers generally had students complete the in-class portion of the daily *Saxon* worksheet in class. While one treatment teacher assigned the homework portion of the worksheet as homework, the other teacher did not use the *Saxon* worksheet exactly as prescribed in the program. She noted that the homework portion of the worksheet was sometimes completed in class rather than independently at home. She did this as she felt her students were often overwhelmed with the amount of homework they were assigned and the *Saxon* Phonics worksheet added too much.

Overall, the teachers had a good hold of the lesson activities but sometimes neglected thorough classroom practice (e.g. Kid Card games, decoding skills with readers) due to time constraints.

Both treatment teachers had the *Saxon* wall cards and alphabet cards displayed in their classrooms. One of the teachers had the alpha strips attached to each students' desk while the other had students keep them in their desks and take them out during phonics instruction. Teachers struggled somewhat initially with the pacing of the program, but felt that the training provided had been helpful to them in selecting the most vital aspects of the program to incorporate. The two treatment teachers were able to keep the pace of 4-5 *Saxon* lessons per week. As mentioned above, one element of the program that teachers did not use regularly was the Kid Card games/classroom practice game activities. The teachers noted that while they enjoyed using the games, time did not permit it regularly and the classroom practice portion of the lesson caused the pace to lag.

Teachers used the spelling component of the program regularly, indicating that the program was incorporated throughout the week. Teachers assigned the spelling sheets to their classes and performed assessments regularly on Fridays.

Highlights –For the most part, the teachers at School B are equally skilled, providing instruction that is purposeful and engaging for most students. However, as is to be expected, style of teaching varies from teacher to teacher to best meet the needs of their group of students. Because each reading group was so different, teachers used different teaching methods to engage their students. In addition, school B began using pieces of the *Saxon Phonics and Spelling* program last year to supplement reading instruction. Therefore, some students had been exposed to the program in the prior year. Prior exposure was examined and results are presented in the Technical Appendix E.

Elementary School C

About the School – Elementary school C is located in a small town in northern Indiana. Surrounded by farmland, the rural community is small and full of school-pride. The school itself opened in the 1950s, currently houses grades K-12 and has been modernized, remaining in good condition. Grades K-6 operate out of a separate wing of the school building away from the middle and upper grades, with a separate entrance and office. A new high school is in the process of being built and is scheduled to open sometime after the start of the school year in the Fall of 2007 which will allow the current building to service grades K-8. Enrollment for grades K-6 was 515 during the 2006-2007 school year, with a student-teacher ratio of 19.

Results from the 2007 ISTEP+ (Indiana Statewide Testing for Educational Progress Plus) show that 79% of third graders achieved a passing score in English/Language Arts, as compared to the statewide average of 74%. Additionally, this school made adequate yearly progress in 2006 for the fifth consecutive year.

The student population is almost entirely white:

- 99% White;
- <1% African American;

- <1% Hispanic.

Approximately 22% of the students were eligible for free or reduced-priced lunches and no students were classified as Limited English Proficiency.

About the Participants – Eight teachers participated in the study: four first grade classes and four second grade classes. Two treatment classrooms and two control classrooms participated at each grade level with approximately 175 students, and the average class size was 22 with a range of 21 to 25.

Reading and language arts instructional periods ranged in length from 1.5 to 2.5 hours, which included *Saxon* instruction for the treatment classes. Treatment teachers reported spending anywhere from 40 to 75 minutes per day on *Saxon* Phonics lessons. Although the timing of the reading/language arts block was flexible across classes, most teachers generally chose to teach reading and language arts in the morning, and all noted that due to the flexible scheduling they would conclude any unfinished morning reading or language arts lessons in the afternoon. While each teacher generally developed their own daily class schedule, all *Saxon* lessons were conducted in the mornings amidst the language arts block.

All teachers reported classes that were mixed in terms of student ability. Student engagement was high in every classroom, and few behavior problems existed. One first grade treatment teacher noted that her students had relatively little home support, with several one-parent families, resulting in a lower homework completion rate and more classroom management issues. Additionally, of the eight teachers at the first and second grade level, the majority (N=5) had been teaching for six years or less, with several in their first or second year of teaching.

The school recently received a large grant from a local pharmaceutical company to improve reading in early elementary grade levels so that 95% of all students would be at or above grade level by 2014. Furthermore, the principal had emphasized phonics as an instructional focus and the school had been in the process of trying to teach teachers phonics-based reading methods. It should be noted that because preschool does not exist in this county, students entering kindergarten have typically been below average, and the school began using pieces of the *Saxon Phonics and Spelling* program last year to supplement reading instruction at the kindergarten and first grade levels. Therefore, some students had been exposed to the program in the prior year. Prior exposure was examined and results presented in the Technical Appendix E.

School C did not begin implementing the *Saxon Phonics and Spelling* program until the end of October, nearly two months after the start of the school year. Several treatment teachers noted the difficulty of this transition in their phonics instruction, having to re-teach old material in an effort to get students accustomed to the terminology and coursework of the new program.

About the Reading and Language Arts Curriculum and Resources – Teachers instructed reading and language arts with a basal program (2001). Several teachers noted their disapproval with the selected program due to its limiting exposure to reading selections and mentioned that the school had selected a new text to replace the basal for the 2007-2008 school year. Despite this, all teachers used the basal almost exclusively for planning purposes, as district guidelines

and an emphasis on phonics drove instruction at both grade levels. First grade teachers would plan together and were generally on the same pace, covering skills at the same time and following the basal sequentially. First grade treatment teachers coordinated weekly and used the same spelling and vocabulary words. Second grade teachers would not plan together, but noted they "touched base" with one another regularly, followed the basal sequentially, and adhered to the district guidelines. Additionally, second grade treatment teachers had team discussions for *Saxon* planning.

In addition to the basal, several teachers also used supplemental materials to augment instructional resources and target specific student needs. Specifically, both second grade control teachers employed the use of a published phonics program (1991 & 1995) to assist with phonics instruction, and one second grade treatment teacher continued to use on occasion a phonics program (1988) she had begun the year with. All teachers referenced the various teacher-created materials they used as well. The two second grade control teachers also used the Four Block Planning Method to help structure their language arts periods. It consisted of four equally important areas of focus: the word block, guided reading block, writing block, and self-selected reading block. In each of the four categories, background knowledge was activated, and a purpose was set for reading.

Several school-wide reading initiatives were in place to support student reading efforts at School C. This included a push for reading such as Bingo for Books, in which students would play bingo to win books, and Book-It, a locally-sponsored program that included prizes and gift certificates to restaurants. The school also implemented Reading Counts, a computer based program similar to Accelerated Reader. Lastly, the Reading Specialist would use a remedial program regularly with remediation students. As a skills-based computer program that would pre-test students and then move them forward progressively, the remedial program was used once a week for 30 minutes, was aligned with state standards, and was also accessible online for students with internet access at home and thus could be used more frequently if desired.

About the Reading and Language Arts Classrooms – All treatment and control classrooms were adequate in space and had sufficient resources. While each classroom had a computer available for teacher use, student computers were not present in classrooms. The school had a computer lab for student use. Classroom libraries were present and well-organized.

A typical language arts lesson at the first grade level consisted of either small group reading instruction or whole group lessons using the basal. In small reading groups, the teacher would spend about 20 minutes with each group of students using a combination of reading strategies including reading aloud to students and having students independently read the text aloud, prompting students to think about the reading material and questioning them about what they had read. Students not in the small group were either reading silently, partner reading, or completing an activity related to the content of the small group lesson such as a story map. Whole group instruction usually began with an introduction to explain the focus of the lesson, followed by a reading of the basal text. The teacher would then discuss the vocabulary words and phonics skills to be practiced, and the class would continue reading, practicing both independently and with partners if time allowed. Depending on the day, students would also engage in workbook activities or use leveled readers for additional practice.

Based on observations, a typical lesson at the second grade level began with an introduction of the skill being taught, in which the teacher would model and the whole class would practice with echoing or coding activities in small groups or as individual work. Skills would be reinforced with group activities such as flash card games or writing practice. Students would then use texts such as poems to find similar words, and the teacher would explain new vocabulary for students to practice through shared writing samples.

While there was no school-wide policy, homework was assigned at least 4 nights a week at both grade levels although the amount varied depending on the teacher. Assignments ranged anywhere from 20 to 30 minutes a night and consisted of reading to parents, writing sentences using weekly words, and studying for spelling or vocabulary assessments. Vocabulary tests were given each Friday on the high-frequency Dolch words, and students were also given fluency assessments regularly.

About the *Saxon Phonics and Spelling* Classrooms – As previously noted, treatment teachers at both the first and second grade levels used the *Saxon* program in addition to their basal phonics component and used it within their daily reading and language arts instructional block. Treatment teachers would coordinate with their grade level counterpart for *Saxon* instruction and were for the most part in synch with one another. During interviews, each treatment teacher commented that time was the most limiting factor in determining which components of the program were regularly implemented.

Implementation at the first grade level varied slightly between the two teachers. One teacher followed the Implementation Guidelines almost exactly as stated, beginning with the introduction to the lesson and Deck Card review which took no more than five minutes. The new increment was introduced with spelling rules, taking another 15 minutes, and the next 30 to 40 minutes were spent on application and continual review. Students would actively and eagerly participate in the board work, which took approximately 15 minutes, and then the worksheet was completed independently and corrected in class. While the letter tiles, handwriting component, and Kid Cards were omitted almost entirely, this teacher did use the *Saxon* spelling component as opposed to the basal. The second first grade treatment teacher, on the other hand, was instructed during a training session to modify the *Saxon* lesson based on what worked, and was therefore less stringent in adhering to the instructions as provided in the teacher's implementation guidelines. The alphabet activity was completed for the lesson warm-up, taking approximately five minutes, although the Deck Cards were skipped. All board work was completed prior to handing out the worksheet. Almost half an hour was then spent on the worksheet, split between both group work on the review section and then independent work on new material, and the worksheet was put in the students' folders to go home at the end of the lesson. This teacher noted that although the letter tiles, Kid Cards and reference books were not used, the decodable readers were implemented midway through the year to great success, and the handwriting component was used to support the handwriting instruction from the basal.

Saxon implementation at the second grade level was more balanced between the two teachers. A typical lesson would begin with a review of the previous night's homework or a game such as coding bingo to engage the students, and then the teacher would complete the lesson warm-up

with an alphabet activity, taking approximately 10 to 15 minutes. The new increment was then introduced, and students and teachers alike would actively reference existing rule charts posted around the room, taking another 10 minutes. The lesson would then either continue as described in the Teacher's Manual, progressing from board work to the worksheet, or the teacher would introduce a game to reinforce new skills and spelling patterns. In both cases, the remaining sections of the worksheet were routinely assigned as homework. One teacher would routinely use the spelling portion, while the other did not do the *Saxon Spelling* but sometimes took the tests just for fun. Neither teacher used the English language history section at the beginning of the lessons as they felt the students did not understand them, nor did they employ the handwriting component or Kid Cards during class. One teacher tutored students after school and would use the Kid Cards and fluency readers for remediation purposes only.

In addition to classroom instruction in the treatment classes, the *Saxon Phonics and Spelling* program was also used by the site liaison/Reading Specialist as a remediation tool with pull-out students from the treatment classes four times a week for approximately 30 minutes. Providing instruction that was two weeks behind the classroom teachers, the *Saxon* lessons were re-taught using the word cards and lesson plans, the worksheet was occasionally redone, and the letter cards and sight words were posted around the remediation room. This instructor was careful to point out that *Saxon* was *not* used as a remediation tool with control classroom students; furthermore, as a result of *Saxon* instruction in the classrooms and as a remediation tool, the number of students requiring additional reading help had decreased from 16 to 6 among the four treatment classes between the fall and spring.

Highlights – School C was extremely eager to begin implementing the *Saxon Phonics and Spelling* program because of their previous successes using *Saxon* programs. Despite many of the teachers being in their first few years of classroom instruction, all teachers were equally capable of engaging their students in purposeful work, and both the first and second grade teaching teams were committed to providing all students with a meaningful learning experience.

Elementary School D

About the School – Elementary school D is located outside of a large city in Texas in a lower-middle class suburban neighborhood. The school itself is relatively new, built in the early 1990s, and is still in excellent condition. Total enrollment for the 2006-2007 school year was 680, and the average class size was 22.

Results from the 2006 TAKS (Texas Assessment of Knowledge and Skills) show that 97% of third graders met or exceeded standards in reading, as compared to the statewide average of 89%. Additionally, this school made adequate yearly progress in 2005-2006, and was rated "Recognized" by the Texas Education Agency based on state test results and dropout rates.

The student population is largely Hispanic:

- 93% Hispanic;

- 6% White;
- <1% African-American;
- <1% American-Indian;
- <1% Asian/Pacific Islander.

Approximately 88% of the students were eligible for free or reduced-priced lunch and 49% were classified as Limited English Proficiency.

About the Participants – Two first grade classes and two second grade classes from elementary school A participated in the study, with one control class and one treatment class at each grade level. There were approximately 77 students and 4 teachers across the classes who participated in the RCT. The average class size was 22, with an observed range from 15 to 23.

The two first grade teachers devoted 2.5 hours a day to reading and language arts instruction, which included *Saxon* instruction for the treatment class. The two second grade teachers, on the other hand, typically spent only 1.5 (control teacher) to 2 hours (treatment teacher) on those subjects. Although the timing of the reading/language arts block was flexible across classes, most teachers generally chose to teach reading and language arts in the morning, and all teachers noted that due to the flexible scheduling they would conclude any unfinished morning reading or language arts lessons in the afternoon. *Saxon* lessons were conducted in the mornings during the language arts block in the first grade treatment classroom and in the afternoon following lunch in the second grade treatment classroom.

All participating teachers described their classes as having a fairly broad range of abilities in terms of performance and average in comparison to other classes at their grade level. The first grade treatment teacher indicated at the end of the school year that she likely had more high performing students than the other first grade teacher, but that she would still classify her class as average overall.

School D reported a mobility rate of 26% for the 2006-2007 school year, and participating teachers noted that this high population of transient families had a direct impact on classroom culture with several changes in roster and the restructuring of classes a few months into the school year. As a result, the second grade treatment teacher was located in a makeshift classroom in the middle of the second grade pod, using removable partitions for walls and a whiteboard on wheels.

Student engagement across the four participating classrooms was high, and most teachers did an excellent job with classroom management, resulting in little to no behavior issues. However, the second grade treatment teacher was a first year teacher with less practice controlling student behavior and engaging difficult students, and because of this, the students in this class were noticeably less engaged and off-task, and the teacher subsequently spent a large portion of time with behavior issues.

School D did not begin implementing the *Saxon Phonics and Spelling* program until the end of October, nearly two months after the start of the school year. Both treatment teachers noted the

difficulty of the transition, having to reteach old material in an effort to get students accustomed to the terminology and coursework of the *Saxon Phonics and Spelling* program.

About the Reading and Language Arts Curriculum and Resources – Teachers instructed reading and language arts with a basal program (2001). Several teachers noted their disapproval with the selected program, citing too many teacher's guides (four), a weak phonics component, and stories that are not integrated well with the rest of the curriculum. Despite this, all teachers used the basal almost exclusively for instruction and planning purposes as it coincides with the district mandated pacing guideline which dictates the instructional objectives that should be targeted each week in reading and language arts instruction. In addition, each grade level is provided with a weekly team planning period that, within the framework of the district pacing guide, helps to keep all classes within a grade level together. The first grade teachers utilized this planning tool to its fullest extent, teaching the same basal lessons on the same day, whereas the two second grade teachers tended to touch base about planning and instruction rather than cooperatively map out their lessons.

While all teachers were compliant with the district pacing guidelines and therefore used the basal almost exclusively, several teachers did use supplemental materials to target specific student needs. Included in the supplementals was a phonics program (1999) used in the first grade control classroom. All teachers referenced the various teacher-created materials they used as well.

Several school-wide reading initiatives were in place to support student reading efforts at School D. Chief among them was D.E.A.R. (Drop Everything And Read), although it was used to varying degrees by the participating teachers. Some referred to it as SSR (sustained silent reading), while another used it to have students read and test using the Accelerated Reader computer program. All four participating teachers agreed that they had their students read for at least 15 minutes every day as part of the district's reading initiative. Additionally, the HOSTS (Helping Other Students To Succeed) program was in place to help at-risk students get additional support with reading skills as members of the community would come in several times a week (usually 3-4) to read with selected students. Lastly, the dyslexia teacher used a computer-based phonics supplement as a pull-out tool with students in need of additional help.

About the Reading and Language Arts Classrooms – As previously noted, the second grade treatment teacher instructed students in a makeshift classroom following the restructuring of the second grade classes. Because of this, the teacher often had to compete with many irregular distractions caused by human traffic and neighboring classrooms for the students' attentions. Additionally, the wall space was severely limited, and many instructional posters were crudely taped to temporary partitions and windows, and the desks were somewhat cramped within the confined space. The teacher admitted that, despite the original imposition, the class had become accustomed to its surroundings and managed to progress fairly well over the course of the year. It should be noted that this teacher was also in her first full year of classroom instruction.

The other participating classrooms were all adequate in space. All treatment and control classrooms had at least one computer for student use, and provided sufficient materials for

students. Classroom libraries were limited, and at the time of observation seemed to be organized haphazardly.

A typical basal lesson at the first grade level began with a brief introduction to the upcoming lesson, using whole group instruction and active student participation as an engagement for the language activity. This generally took about 10 minutes. The teacher would then read a short story or section of a story aloud to the class as whole group practice for the language activity, followed by students reading independently to practice on their own, which took another 15 to 20 minutes. This independent practice was reinforced by the completion of a workbook page on the appropriate language activity. The remaining class time was used to have groups of students read together, either in leveled groups with the teacher while the rest of the class read independently, or as groups in the students' seating arrangement.

Based on observations, typical instruction at the second grade level varied between the control and treatment teacher. The control teacher would typically begin the morning reading/language arts period with board work, which took approximately 15 minutes to review and correct as a class. The teacher would then use the basal to introduce the lesson topic and new information. Whole group practice would occur as the teacher read aloud to the class and had students participate with the language activity, and was followed by independent practice as the class would either read together and discuss or play a game for reinforcement. The treatment teacher, in contrast, began with an oral review of a previous lesson or skill and had students verbally participate to demonstrate understanding. Depending on student engagement, this could take up to 15 minutes. The teacher then read a text aloud to the students, stopping to clarify and check for understanding. Students would then partner read or work independently in a writing assignment related to the content of the lesson.

At both the first and second grade level, district benchmark assessments were given every 9 weeks, which teachers reported was after approximately 4 to 5 stories in the basal. In addition, the second grade administered brief 10 minute basal tests every 2 weeks. As a school, students were tested twice a year with the ITBS.

Homework was assigned every night at both grade levels, although the amount varied depending on the teacher. All participating teachers noted that there was very little home support, which affected students' efforts at home. Homework assignments ranged anywhere from 10 to 30 minutes a night, comprised of phonics, spelling and reading. Both treatment teachers reported assigning *Saxon* homework very rarely, as the work was usually attended to in class.

About the *Saxon Phonics and Spelling* Classrooms – As indicated, both treatment teachers used the *Saxon* program in addition to their basal phonics component, and used it within the 2½ hour reading and language arts instructional period. Pacing issues existed, and there was a notable difference in implementation between the two treatment teachers.

The first grade treatment teacher followed the Teacher's Manual almost exactly as outlined each lesson. Spending close to 60 minutes a day, a typical lesson began with a full warm-up which consisted of the introduction of the lesson objective, a brief language activity, and a quick review of the Deck Cards in random order, which would take no more than 10 minutes. The new

increment was introduced, and students would listen and participate as the board work was then completed, usually taking another 20 to 25 minutes. The remainder of class was then divided between the worksheet and classroom practice, depending on the lesson. The teacher noted that the lessons became increasingly longer and more complex as the year went on and skills built upon one another, and because of this the integrity of lesson completion was compromised somewhat due to time constraints. Students would consistently use the decodable readers and the letter tiles when lessons called for them, and to a lesser degree the teacher would use the fluency readers. The spelling portion was taught as time permitted, although the site word evaluations, handwriting instruction and fluency assessments were not used.

The second grade treatment teacher admittedly spent only 30-40 minutes on *Saxon* each day. As a first year teacher teaching in a makeshift classroom, student behavior played a significant role in time management and lesson completion. Although this teacher paired with the first grade treatment teacher for planning after struggling with pacing issues, portions of the treatment curriculum and certain lesson segments were regularly omitted. A typical lesson began with a brief, 10 minute warm-up, during which time students engaged in a language or alphabet activity and the teacher quickly reviewed a shuffled and shortened Deck Card set. The new increment was then introduced, including the new Keyword or Wall Card and reference to existing posters around the classroom, generally taking the next 10 minutes. After this, students would actively participate in the board work, but classroom behavior would then disrupt the lesson. In general, the spelling portion was not consistently used, the handwriting portion was discontinued due to time constraints, and the fluency and decodable readers were used more regularly as supplemental materials, sent home for homework, or skipped altogether as opposed to including them in the *Saxon* Phonics lesson.

While there was little consistency between the two treatment teachers as far as lesson implementation, both teachers had the *Saxon* posters referencing appropriate posture up and indicated which student pose they expected at various times during the day. Additionally, both teachers mentioned that they rarely included the Kid Card games, mostly because of time, and neither used the running records regularly.

Highlights – Instructional techniques varied rather significantly between the teachers at School A, although it was evident that all teachers did their best to engage their students in purposeful work. Because of the district guidelines, the two first grade teachers were on pace with each other, as were the two second grade teachers. However, pacing issues certainly arose for treatment teachers, and the second grade treatment teacher's struggles with implementation were more significant than those of the first grade treatment teacher.

Elementary School E

About the School – Elementary school E is located outside of a large city in Texas in a lower-middle class suburban neighborhood. The school building is 50 years old but has been renovated several times, modernized recently, and remains in excellent condition. Total enrollment for the 2006-2007 school year was 568, and the average class size was 19.

Results from the 2006 TAKS (Texas Assessment of Knowledge and Skills) show that 92% of third graders met or exceeded standards in reading, as compared to the statewide average of 89%. Additionally, this school made adequate yearly progress in 2005-2006, and was rated "Recognized" by the Texas Education Agency based on state test results and dropout rates.

The student population is largely Hispanic:

- 85% Hispanic;
- 8% White;
- 6% African-American;
- <1% American-Indian;
- <1% Asian/Pacific Islander.

Approximately 92% of the students were eligible for free or reduced-priced lunches and 42% were classified as Limited English Proficiency.

About the Participants – Six classes participated in the *Saxon Phonics and Spelling* study, with two treatment and one control classroom at both the first and second grade level. There were approximately 92 students and 6 teachers across the classes who participated in the RCT. The average class size was 19, with an observed range from 10 to 20.

All teachers devoted 2.5 hours a day to reading and language arts instruction, which included *Saxon* instruction for the treatment classes. Although the timing of the reading/language arts block was flexible across classes, most teachers generally chose to teach reading and language arts in the morning, and all noted that due to the flexible scheduling they would conclude any unfinished morning reading or language arts lessons in the afternoon. *Saxon* lessons were conducted in the mornings during the language arts block.

All participating teachers described their classes as having a fairly broad range of abilities in terms of performance and average in comparison to other classes at their grade level. Teachers commented that although classrooms at both grade levels were comparable, overall student performance was the lowest it had been in a number of years. Student engagement across the six participating classrooms was high. It was apparent that where behavior issues had existed at the beginning of the year, teachers had found in-class solutions or the administration handled, resulting in little to no classroom management problems.

School E did not begin implementing the *Saxon Phonics and Spelling* program until the end of October, a full two months after the start of the school year. Treatment teachers noted the difficulty of the late transition, getting students accustomed to the terminology and coursework of the new program, as well as having to deal with pieces of the new curriculum. These treatment teachers admitted to excluding certain *Saxon* components that they felt they could've incorporated had they started at the beginning of the year, but due to the lack of planning ahead of time, did not.

About the Reading and Language Arts Curriculum and Resources – Teachers had at their disposal a basal (2001) for reading and language arts instruction, although several commented that they rarely relied on this as a planning device or instructional tool. Instead, the basal was used more as a supplemental resource for writing and grammar practice and in the case of the 1st grade treatment teacher, was not used at all. Guided reading was instituted as the main initiative for improving students' reading skills, in conjunction with the district mandated pacing guideline which dictates the instructional objectives that should be targeted each week in reading and language arts instruction. At the first grade level, the team leader (1st grade treatment) did all of the lesson planning for the teachers, helping to ensure equal pacing. While other first grade teachers may have altered the plans somewhat based on students' understanding, all three participating teachers agreed they were together in their instruction. All three participating second grade teachers would coordinate in terms of reading and language arts planning, but two (control and treatment) would synchronize their instruction on a daily basis, while the third (treatment) would plan separately and use different instructional techniques. Still, pacing and content taught was similar across all 2nd grade classes.

All teachers were compliant with the district pacing guidelines, and while many used the basal, they also supplemented extensively to target specific student needs. Specifically, teachers would use leveled readers for students of differing ability levels as well as trade books for their reading groups. In addition, one second grade treatment teacher relied heavily on the Four Blocks program for language arts planning, and used a series of blackline masters for classroom worksheets. All teachers referenced the various teacher-created materials they used as well. In the treatment classrooms, the *Saxon Spelling* component was implemented initially, but as the school year progressed, it was dropped in favor of the district mandated spelling lists.

Several school-wide reading initiatives were in place to support student reading efforts at School E as part of the district reading initiative. While guided reading was again mentioned, D.E.A.R. (Drop Everything And Read) was another program in place, although it was used to varying degrees by the participating teachers. Some teachers used it when assigned class work was completed, others used it after lunch for approximately 15 minutes, while another used it to have students read and test using the Accelerated Reader computer program up to 45 minutes each day. All participating teachers agreed that they had their students read for at least 15 minutes every day as part of the district reading initiative. Additionally, the HOSTS (Helping Other Students To Succeed) program was in place to help at-risk students get additional support with reading skills as members of the community would come in several times a week (usually 3-4) to read with selected students.

About the Reading and Language Arts Classrooms – All treatment and control classrooms were adequate in space and had sufficient resources. Each had at least one computer for student use, and teachers reported taking their students to the computer lab on a regular basis. Classroom libraries were present and well-organized.

A typical reading lesson at the first grade level consisted of either small group work or centers in which there were a variety of different activities including art, social studies, spelling, reading, worksheets, math flash cards and poetry composition. Students would rotate through the centers completing the assignments while the teacher conducted a guided reading session with a small

group. This session usually started with the recitation of sight words, a discussion of reading strategies and a review of vocabulary words. The teacher would then lead students through a picture walk and conducted a general discussion of the story. Students took turns reading out-loud, and the teacher would occasionally stop to ask comprehension questions and focus on individual student fluency and intonation.

A typical reading lesson at the second grade level also consisted of small group work, rotating centers, and guided reading groups. Teachers had in place at least two centers, one of which was a teacher led guided reading session where teachers and students would take turns reading passages out loud and then the teacher would guide discussion focusing on new vocabulary, comprehension and reading fluency. The remainder of students would usually work in small groups at their desks, focusing on spelling, grammar and handwriting activities, and would read independently when assignments were completed.

At both the first and second grade level, district benchmark assessments were given every 9 weeks. As a school, students were tested twice a year with the ITBS. Homework was assigned every night at both grade levels, although the amount varied depending on the teacher. All participating teachers noted that there was very little home support, which effected students' efforts after school. Homework assignments ranged anywhere from 10 to 30 minutes a night, comprised of phonics, spelling, reading and math. Treatment teachers reported assigning *Saxon* homework very rarely, as the work was usually attended to in class.

About the *Saxon Phonics and Spelling* Classrooms – As indicated, all treatment teachers used the *Saxon* program within the 2 ½ hour reading and language arts instructional period each morning. While pacing problems were not specified as a major issue, the two treatment teachers at the first grade level were not instructing the same lesson at the same time, nor were the treatment teachers at the second grade level. Despite this, instructional strategies within the lesson were similar, and all teachers followed the *Saxon* lessons *mostly* as outlined in the Teacher's Guide, making slight modifications based on time constraints.

First grade treatment teachers both spent approximately 45 minutes a day instructing *Saxon* lessons, and commented that they felt their students lost interest if they spent an hour or more on phonics. Additionally, during the lesson warm-up activities, which were composed of a language/alphabet activity and the Deck Card review and generally took 15 minutes, both teachers ensured their students engaged not only their minds but their bodies as well by using calisthenics in their instruction. For example, students would jump as they recited old sight words or letter sounds. The new increment was then introduced, and the classes would proceed to the application and continual review of the material. One first grade treatment teacher placed a greater emphasis on completing the worksheet practice and would have students complete the in-class and homework side in class. The other teacher followed the implementation guidelines closely and would complete the board work section in its entirety before having students return to their desks to begin individual practice with the worksheet. Between the board work and the worksheet, each teacher spent almost 30 minutes on the application and review. The remaining time during the *Saxon* lesson, usually 15 minutes, was spent on classroom practice using the fluency and decodable readers to enhance fluency skills or find words that pertained to the lesson.

The second grade treatment teachers followed the *Saxon* plans more closely as detailed in the Teacher's Guide and placed a greater emphasis on precision with the Deck Card review at the beginning of the lesson. Each teacher spent at least 10 minutes on the warm-up activities including specific identification of the lesson's objective. Both teachers also introduced the new *Saxon* card during the teaching of the new increment and referred students to posted rule charts, spelling charts, and *Saxon* keyword charts around the classrooms. The remaining time in class, anywhere from 15 to 30 minutes, was spent on the application and review of the material. One teacher had students actively participate in the board work and then had students complete the worksheet independently at their desks. The other second grade treatment teacher omitted the board work section of the review, spending more time on helping students individually with the worksheet during class. Neither second grade treatment teacher had students participate in classroom practice during the spring observation.

All four treatment teachers at School E used the *Saxon* program exclusively for phonics instruction. While the exact implementation varied slightly from classroom to classroom, most teachers did not use the handwriting instruction or the Kid Cards except for remediation purposes. As previously noted, the spelling portion was not used regularly because of the district spelling lists in place. All teachers mentioned that they tried to have students complete the homework portion of the worksheet in class; otherwise it would not be finished by the students. During interviews, each treatment teacher commented that time was the most limiting factor in determining which components of the program were regularly implemented.

Highlights – Although teaching styles varied between classrooms, all teachers at School E were equally skilled and attempted to continually invest their students in purposeful work. District guidelines and communication between teachers at each grade level helped to keep pacing a non-issue, as teachers did their best to meet the needs of their students.

Elementary School F

About the school – Elementary school F is a large, modern facility built 6 years ago and is located in urban Georgia. The school is in excellent condition and has more than adequate space including a large gymnasium, library, separate art rooms, music rooms, science labs and computer labs. Enrollment during the 2006-2007 school year was 563.

The school's population is largely African American:

- 3.8% White
- 8.2% Hispanic
- 81.3% African American
- 1.1% Asian
- 5.5% other

Approximately 15% of students spoke a language other than English as their primary language. The school also contained a high free/ reduced population, with 83% of students eligible for free/reduced lunch. Breakfast is served from 7:30 to 8:00 every morning for students at school F.

About the participants - There were five first grade classes participating in the study (two treatment and three control) and five second grade classes in the study (two treatment and three control). The average class size across all participating study classes was 19, ranging from 18 to 21 students per class. Each class, at both first and second grade, had at least a two hour reading and language arts block every day. The first thing that took place each day within the block was DEAR time (Drop Everything And Read) for 20 minutes. Also within this block, *Saxon Phonics was Spelling* instruction was given. Depending on the day, this block may also have included compulsory Spanish instruction for approximately 30 minutes. The Spanish lesson was taught by a roaming Spanish Language Instructor. Spanish Instruction was compulsory for grades 1-5 for 30 minutes each day.

Most teachers described their classes as having a broad ability range and average overall in terms of performance in comparison to the other classes. One first grade treatment teacher, however, described her class as "higher performing" overall, but still containing a broad range of abilities. While some moderate student engagement and behavior issues existed, these were equal across all classes.

About the Reading/LA Curriculum and Resources – The same basal program was used for all teachers at both first and second grade (2003). The basal program is designed to provide a lesson plan for each day of the week using literature to tie in all the parts of each lesson. Typically, Day 1 consists of introducing the vocabulary, looking at and reading the literature passage that goes along with the lesson. Day 2 consists of group work, pair reading or an activity practicing the skill emphasized in the lesson. The Day 3 lesson is typically some type of writing to reinforce the concept and may also involve rereading the literature. Day 4 involved some wrap-up of writing, and using the book to review the story focusing on comprehension and vocabulary. Day 5 was usually a wrap-up activity or center activities with review of the week's lesson.

There is substantial commonality among the teachers at school F in terms of what is covered in the basal and the skills that are emphasized each week. Each week, 1st and 2nd grade teachers get together to determine which stories and content they are going to cover in the basal and to determine which skills are going to be addressed. While the skills being emphasized as part of the core language arts curriculum is common across the teachers, there is some flexibility for teachers in terms of how they choose to address such skills. For first grade, the teachers all had a common planning period once per week where they discussed what would be covered in the upcoming week. A couple of exceptions were that one of the first grade control teachers was often one story behind the group due to the pacing her class required. Additionally, one of the first grade treatment teachers often chose to use her own selected leveled reader to teach the skills for the week as opposed to using the story from the basal. So while she was teaching the same skills, she sometimes chose to use a different story.

The second grade teachers also met each week to determine which skill would be emphasized. The second grade teachers used the basal as a guide, but several of the teachers noted that they liked to pull in more outside resources to supplement their lessons. Control teachers tended to supplement the basal with more outside resources than the treatment group teachers. They used a variety of resources including teacher created materials, lots of trade books, leveled readers and materials from other publishers. Several of the 2nd grade teachers also used phonics programs to supplement the basal.

The media specialist at school F encouraged use of Accelerated Reader, a computerized program that tests students on leveled readers, for both 1st and 2nd grade teachers. All teachers except three teachers (one control and two treatment teachers) used Accelerated Reader with their classes. Generally, the content covered and materials used in all classes were very similar. The primary differences between the classes would be individual teaching styles and slight differences in supplemental materials that teachers used in addition to the basal reader.

About the Reading/LA Classrooms- There was more than adequate space in all treatment and control classrooms and resources were plentiful. There were computers available for student use in the individual classrooms and in the library. There was also a computer lab available for classes to use. Independent reading collections were available in each classroom and included a variety of both non-fiction and fiction selections along with leveled readers and a variety of different genres.

As noted previously, each day the reading/language arts block began with a 20 minute Drop Everything and Read (DEAR) period where students read independently or with a partner. Following DEAR, a typical basal lesson varied depending on where teachers were in the 5 day lesson sequence. Most lessons began with a warm-up activity or a review of spelling and/or vocabulary words. Following the warm-up or review, the next 20-30 minutes of each lesson was usually done as whole-group, and included new information being taught, and students participating in work on the board. Then teachers either had students work independently on a writing assignment or a worksheet, or break into groups from small group instruction.

Both 1st and 2nd grade teachers' generally assigned exercises as in-class independent work four to five times per week. Teachers also assigned reading/language arts exercises as homework regularly. Homework assignments most often came from the basal workbooks or teacher created materials. At 1st grade, teachers noted that on average, homework was turned in 90% of the time. The 2nd grade teachers indicated a 95% turn-in rate for homework assignments. For each grade level, the same assessments were administered including state assessments, weekly tests (from the basal) and unit tests.

About the Saxon Phonics and Spelling Classrooms- The *Saxon Phonics and Spelling* program was used during the school's reading/ language art block and was all taught in the mornings in both 1st and 2nd grades. All teachers followed the lessons closely, making slight modifications based on time constraints. Treatment teachers at school F generally followed the implementation guidelines closely and taught the *Saxon Phonics and Spelling* program with great fidelity. However, teachers did note that at the start of the school year they had difficulty integrating

Saxon Phonics and Spelling with their basal program and as a result didn't progress as quickly through the *Saxon* lessons at the start of the year.

After continued questions about the pacing and implementation of the *Saxon Phonics and Spelling* program, PRES Associates worked with Saxon publishers and School F to arrange for the teachers to observe Saxon being taught by teachers at a nearby school and to meet to ask any questions about the program. Teachers found this training tool to be extremely helpful. Once being able to see the program in use they were able to better understand how to fit it into their own classrooms. Two teachers in particular, one first grade and one second grade, were noted as being very engaging and having exceptional classroom management skill during the Saxon lessons.

A typical *Saxon Phonics and Spelling* lesson for both first and second grade began with the lesson warm-up and reviewing the letter/sound decks. One second grade teacher even had his students complete this review to a rap song, which was very engaging to students. Following this review, teachers taught the new increment of the lesson and introduced new cards as the lesson called for. Then students engaged in board work to complete examples of the new increment that was taught. After board work, the teachers passed out the worksheet and students worked independently to complete it. When students were finished teachers did a quick check of the worksheet before the second side was assigned as homework. It should be noted that one of the first grade teachers, while completing all of the lesson components, often rearranged the order. For example, she would teach the new increment and then break up the lesson with the review of the deck cards.

The first grade teachers were using the spelling component of the program. However, one teacher noted that while she integrated the Saxon spelling words, she still relied heavily on the basal for spelling instruction. Neither of the second grade teachers were consistently using the spelling portion of the Saxon Phonics and Spelling program. Additionally, the teachers did not always use the Kid Cards and other classroom practice activities due to time constraints. Toward the middle to end of the school year, as teachers became more comfortable with the program, they used these resources more often. Other than the variation of use noted here, all teachers were using the program as prescribed and implementing effectively.

Highlights: School F possessed a positive school climate that appeared to emphasize student learning. Teachers across treatment and control groups were all committed and effective. It was noted during observation that two of the treatment teachers were particularly engaging in their instructional delivery. Though there were varying teacher styles, the level of experience and level of competence seemed equal across both groups. Initially, there were some concerns about the treatment teachers pacing of the *Saxon Phonics and Spelling* program. Following additional training and observations of other teachers using the program, such problems were resolved. It is noteworthy that the treatment teachers found it so useful to observe the *Saxon Phonics and Spelling* program being used by teachers in another school. Several of the treatment teachers commented that this really helped them to visualize and get ideas as to how to effectively implement the program. In fact, some mentioned that they thought a more effective training format would have been to let them observe the program being used first, before any orientation training.

Teachers in school were all piloting new Promethian boards this year. Teachers were provided with training and commented that it had given them more creative ways to visually present materials and skills. Virtually all the teachers observed used the Promethean Boards regularly and effectively in their instruction. They also commented that it helped with keeping the kids focused, though there was a learning curve associated with using it for the first time.

Appendix C:

Teacher Ratings of *Saxon Phonics & Spelling* Program and Use of Resources

Table C1. Percent and Rating Given to Usefulness of *Saxon Phonics and Spelling* Resource

	Percent who indicated useful to very useful	Mean*	Std. Deviation
Review Decks	93.3%	4.67	.617
Teacher's Manual	93.3%	4.60	.632
The program as a whole	86.7%	4.40	.737
Teacher's Resource Binder	80.0%	4.33	.816
Wall Cards/Posters	73.3%	4.27	.884
Student Decodable Readers	78.6%	4.14	1.231
The *Saxon* Phonics daily worksheet	73.3%	4.13	.834
Assessments provided with program.	80.0%	4.13	.743
Fluency Readers	69.2%	3.85	1.068
Spelling Dictionary and Reference Booklet	57.1%	3.79	1.122
The Language/Alphabet Activity from the Teacher's Manual	64.3%	3.79	1.051
Kid Card Games	64.3%	3.79	1.122
Fluency Instruction Booklets/Masters	54.5%	3.64	1.286
Handwriting Instruction Booklets and Masters	75.0%	3.63	1.408
Alphabet Handwriting Strips	60.0%	3.60	1.242
Letter Tiles	36.4%	3.27	1.348

*Based on scale of 1-not at all useful to 5-very useful.

Table C2 Percent Use (and Mean) of *Saxon Phonics and Spelling* Componen

Saxon Phonics and Spelling Component	Mean*	Percent who indicated often (2-3 times/week) to everyday
Teach new increment	4.80	95.3%
Assign Worksheet for students to complete independently	4.75	95.3%
Complete board work	4.66	93.8%
Use *Saxon Phonics and Spelling* Keywords	4.59	89.1%
Review letters, sounds, and spelling after warm-up	4.53	87.5%
Introduced the Lesson (including objective)	4.16	84.4%
Used school/home reinforcement	3.92	68.8%
Use the Deckcards/Letter Tiles/Related Wall Cards	3.86	64.1%
Engage in a Language/Alphabet activity	3.67	56.3%
Used Decodable Readers	3.30	45.3%
Use Student Spelling Dictionary and Reference Booklets	3.33	43.8%
Conduct classroom practice with a game or activity	3.06	37.5%
Used Fluency Readers	2.89	31.3%
Used *Saxon Phonics and Spelling* Handwriting Instruction	1.84	15.6%
Fluency Assessment	2.00	7.8%

*Based on scale of 1-not at all used to 5-everyday.

Assessment Use	Average Percent Use (weekly)
Saxon Phonics and Spelling Oral and Written assessment	70.4%
Saxon Phonics and Spelling Site Word Evaluation	65.0%

Appendix D:

Implementation Guidelines

SAXON PHONICS AND SPELLING RCT STUDY
IMPLEMENTATION GUIDELINES FOR TEACHERS

Welcome, and thank you for participating in the *Saxon Phonics and Spelling* study. We believe your experience with our study will be rewarding and enjoyable.

We understand that it may be challenging to alter instructional practices and implement a new supplemental reading program. Therefore, we greatly appreciate the time and effort you will be putting into making this study a success. However, we also realize that there will be obstacles and challenges as you begin to implement this program. Under these circumstances, we want and need to hear from you; we will make every attempt to guide you through those challenges. In fact, it is critical that any problems you encounter be addressed as soon as possible to ensure that this program is being implemented to its full potential. Feel free to contact Dawn Gertsch, Research Coordinator for PRES Associates at 1-866-599-PRES or dgertsch@presassociates.com if you have any questions, problems, concerns and so forth.

The following provides answers to some common questions teachers may have related to this study. Please read through all of these questions/answers. Should you have further questions, please contact PRES Associates.

WHY IS THIS RESEARCH BEING DONE?

As you are aware, the No-Child Left Behind (NCLB) of 2001 requires that educational materials and strategies used by educators in the classroom must be proven by scientific research to improve student achievement in the classroom. *Saxon* has developed a strong research model for determining that their programs are scientifically-based and successful. As part of this ambitious research agenda, *Saxon* has contracted with PRES Associates[44], an external educational research firm, to conduct a rigorous quantitative randomized control trial (RCT) on the effectiveness of the *Saxon Phonics and Spelling* during the 2006-7 school year. The study will contribute to the growing research base behind *Saxon Phonics and Spelling* and the effectiveness of different approaches to reading instruction.

WHY DO I NEED PROFESSIONAL DEVELOPMENT?

It takes more than a good curricular program to raise students' phonics and spelling ability. It also takes good teachers with a thorough understanding of the curriculum and who are supported by professional development, school administrators, and parents/guardians. To this end, it is hoped that through the professional development training session provided by *Saxon* on the use of its *Saxon Phonics and Spelling* program, all "treatment" teachers participating in the study will gain the knowledge and skills to successfully implement this program right from the start.

[44] PRES Associates is an external, independent, educational research firm with an established track record in conducting large-scale, rigorous evaluations on the effectiveness of research materials.

As you will learn, this program offers numerous methods to enhance your basal reading curriculum and improve students' phonics and spelling. In order to implement this program successfully, it is essential that teachers have a thorough understanding of the *Saxon Phonics and Spelling* program. Rather than having teachers figure it out on their own, professional trainers will guide you through this process.

WHY DO I NEEDTO FOLLOW THESE IMPLEMENTATION GUIDELINES?

The Teacher Implementation Guidelines were developed as part of the *Saxon Phonics and Spelling* RCT. The guidelines are designed for "treatment" teachers to use while implementing the new program. The guidelines point out key program components that must be implemented during *Saxon Phonics and Spelling* lessons. These key program components have the greatest influence on student learning and performance and therefore should be implemented. In addition, it is critical to ensure that all "treatment" teachers are using the program with fidelity. That is, if teachers are modifying the program to an extent that it no longer resembles the original program, it affects the validity of the study. In sum, by providing these implementation guidelines, we are attempting to (1) maximize the potential of this supplemental program, and (2) ensure that the program is being implemented with equal fidelity across teachers. To reiterate, it is essential that all "treatment" teachers implement the program fully as prescribed in the following implementation guidelines.

That said, we do not expect that all teachers will teach in the same style or manner, or even use all of the same ancillary program resources. We know that each teacher has different teaching preferences and different student needs. We trust your professional judgment and ask that you try to implement the program as best you possibly can while meeting your instructional needs.

WHAT MATERIALS ARE PROVIDED WITH THE SAXON PHONICS AND SPELLING STUDY?

You have been provided with the following *Saxon Phonics and Spelling* program materials.

- **Teacher's Manual (Volumes 1 and 2 and Resource Binder):** This resource provides clear lesson plans that introduce key phonics strategies and skills. The manuals come in binders that contain 140 individual, easy-to-use lesson booklets. Lessons are carefully sequenced so that each learning increment builds on previously taught skills and concepts, promoting student confidence.

  Additionally, teacher's materials also include a **Teacher's Resource Binder** with blackline masters of phonemic awareness assessments, parent letters, recording forms, and selected student pages. The Teacher's Guide provides you with a detailed description of program components and background on the research base behind the program in the "Instructional Overview" found in the first few pages.

- **Classroom Materials:** The program includes a variety of non-consumable classroom materials that are versatile support materials which help reinforce critical skills and concepts. Classroom materials include:
  - *Four sets of average leveled colorful **Fluency Readers** (26 titles for grades 1 and 2)
  - ***Review Decks** which expand, enhance and integrate each day's instruction by providing practice and review
  - ***Wall cards** serve as reminders for children and reinforce the phonics and reading

concepts they are learning
*__Posters__ to facilitate classroom managements and serve as reference tools to remind students of phonics and concepts and the most effective writing and listening positions
*__Kid Cards__ provide a fun, hands-on way to review and remediate phonics skills
*__Audiocassette tape__ (for teacher use)

- **Student Materials:** The student materials provided as a part of the *Saxon Phonics and Spelling* program allow children to practice reading controlled, decodable text –including previously taught sight words—allowing children to practice and review phonics and fluency skills. The student materials for *Saxon Phonics and Spelling* are all consumable. Student materials include:
  *__Student Decodable Readers__ include engaging stories and illustrations motivate children to read, and the black and white format allows them to color and individualize their own set of books.
  *__Alphabet handwriting strips__ for each child to remind them of alphabetical order, letter sounds and proper letter formation.
  *__Letter tiles__ used to practice spelling, alphabetizing, and letter recognition (1st grade only).
  *Student __Spelling Dictionary and Reference Booklets__ serve as a reference for grade-level words and commonly used spelling patterns.
  *A classroom set of all __student worksheets__ and __assessments__ with __file folders__ to be organized by lesson and stored in stackable, reusable __plastic crates__.

WHAT IS THE INSTRUCTIONAL PHILOSOPHY OF THE SAXON PHONICS AND SPELLING PROGRAM?

Saxon Phonics and Spelling is a success-oriented program that enables most children to develop a solid foundation in phonics and thus become successful readers and spellers. In keeping with the *Saxon* philosophy of incremental development and continual review, the program presents new learning in small increments that children review daily for the entire year. This method of reinforcement gives children the practice they need to achieve success.

The program emphasizes 7 key elements. They are:

- **Controlled Vocabulary and Reading Practice:** A controlled vocabulary is used throughout this program. Children read only those words containing letters/letter clusters, sounds, and syllable division patterns that have been taught. This provides immediate reinforcement of the concepts learned and also ensures that children experience continues success as they learn to read. Although they will gradually transfer their newly acquired skills to other reading material, it is best to focus children's independent reading on the controlled texts provided until they have mastered some basic decoding skills. These texts will gradually increase in length and difficulty.

- **Role of Literature:** Children should be provided numerous opportunities to experience the rhythm of language, to enrich their vocabularies, and to develop a love for reading. For these reasons, reading authentic literature to children is strongly recommended. Literature will introduce them to rich vocabulary. Syntax and themes and will also help

them make a smooth transition from learning and applying decoding skills to reading for meaning and pleasure. The **Annotated Bibliography** is designed to help you find appropriate children's literature to accompany *Saxon Phonics and Spelling* lessons.

- **Coding:** Coding is one tool used to help create successful readers. Children are taught how to code words by marking common vowel patterns and letter clusters, which helps them identify the sound of each letter/letter cluster and thus read the words. This method gives children the ability to approach new words confidently and familiarizes them with dictionary pronunciation and phonetic coding so that they will understand many of the pronunciation symbols used in dictionaries. The overall goal is to teach children how to read; coding is simply a tool that helps them achieve that end.

- **Spelling:** A series of short, simple rules explaining typical spelling patterns will be taught and continually reviewed throughout the program. By practicing the most common ways to spell specific sounds, children can successfully spell all words with regular spelling patterns and are not limited to memorizing words for a spelling test. Each spelling rule is posted on a wall card and listed in the **Student Spelling Dictionary and Reference Booklet.** Words that do not follow the spelling rules, called sight words and irregular spelling words, are also taught, practiced and listed in the reference booklet for quick, easy access.

- **Assessment and Remediation:** Oral and written phonics assessments, sight word evaluations, spelling tests, and optional reading fluency assessments are built into the program. Designed to meet screening, diagnostic, instructional, and evaluative objectives, assessments gauge how well children are learning and retaining concepts and will help you determine whether to adjust the pace of instruction. Assessments should be considered diagnostic tools rather than grading tools; complete them promptly to identify areas for remediation. Classroom and individual assessment forms are available to chart results and to help detect problems.

- **Handwriting:** The letters shown on several components in this program are modeled after the Time Roman typeface which is what most children will encounter when reading. However, if they are to become successful readers, children must understand the correlations between print style and handwriting style. In *Saxon Phonics and Spelling* the teacher determines the choice of handwriting style. Although there may be disagreement about the appropriate style of handwriting to teach to 1st and 2nd graders, this program accommodates most ideologies. To facilitate the teaching of handwriting, a **Handwriting Instruction booklet** and accompanying **handwriting masters** are provided. The booklet provides explicit instruction for teaching children how to write, and the instruction is easily adaptable to handwriting styles other then the one provided on the masters.

- **Reading Fluency:** To become successful, independent readers, children must acquire not only basic decoding skills but also reading fluency. Fluency is the ability to read quickly, accurately, and expressively. It is a necessary skill for developing reading comprehension. Through the use of **leveled fluency readers and fluency masters**, *Saxon Phonics and Spelling* provides explicit, systematic practice for children to develop reading fluency. In addition, the **Fluency Instruction booklets** provides guidance for teaching fluency, including modeling fluent reading, guiding children through repeated

readings, and providing opportunities for independent reading. Instruction is also included for administering assessments and tracking progress. Finally, tips for at-home support are provided.

All of this information can be found in more detail in your Teacher's Manual.

The following pages have the implementation guidelines that we would like you to follow. As you are aware, we will also be monitoring implementation of this program. We will conduct two classroom observations, in part, to determine the extent to which teachers are implementing all key components. In addition, teachers will complete monthly logs to indicate the extent to which they used key and optional components. We will provide more detail on the teacher logs shortly. Together, this data will help us determine the fidelity of implementation.

Thank you very much for your time and your participation in the *Saxon Phonics and Spelling* Study. You are an integral part of this study and we appreciate your time and assistance. Please contact PRES Associates with any questions or concerns. We look forward to working with you.

2006 *SAXON PHONICS AND SPELLING*

IMPLEMENTATION GUIDELINES

Because the *Saxon Phonics and Spelling* program is designed to **supplement** and **enhance** your basal reading program, the program offers teachers flexibility in how they use the program to meet their needs. That said, there are certain key elements of the program that are considered essential to its effective use, and are necessary to use with each lesson.

Please follow these guidelines as you implement the Phonics and Spelling program. All of these items are considered critical to the success of the program.

√ **PACING**. The *Saxon Phonics and Spelling* program is designed to be used 5 days ***per week (four lessons and an assessment).*** It's usually best to teach one lesson per day, using the suggested reinforcement activities to strengthen weak areas. The fifth day of the week (assessment day) may also be used to reteach a difficult lesson, remediate, or meet special school/district objectives. *See page 13 of the Instructional Overview in the Teacher's Manual for tips on adjusting the pace.*

Each lesson (except every fifth lesson) is designed to follow the same instructional path:

1. Lesson Warm-Up
2. New Increment
3. Application and Continual Review
4. Classroom Practice
5. School/Home Reinforcement

* Assessment and Remediation (Assessment lesson only-every fifth lesson)

Use the lesson outlines provided in your **Teacher's Manual** for an overview of each section of the lesson path for each lesson. It is suggested that you spend approximately 1 hour per day for 1st grade and 50 minutes per day for 2nd grade using the *Saxon Phonics and Spelling* program with your class. You may find that you spend more time on each lesson when you first begin using the program. However, as you adjust to the format of the program, you should be within the approximations above.

Please note: Because of the systematic, spiraling approach of the *Saxon Phonics and Spelling* program, it is imperative that you follow the lesson sequence in the order presented by the program.

ACTIVITIES TO BE DONE EACH LESSON (1 PER DAY AND 4 DAYS PER WEEK):

√ **LESSON WARM-UP (APPROXIMATELY 10 MINUTES):** This section of the program is critical. Follow the lesson plan as outlined in the Teacher's Manual. In order to meet this key program component, teachers should:

- **Introduce the Lesson:** You should be sure to introduce the lesson by explaining its **Objective** from the beginning of each lesson. Students should know what they will be learning, how they should learn it, and why.

- **Engage in Activity:** On most days, children engage in a Language/Alphabet activity. These activities grow in complexity throughout the year and include dictionary skills, alphabetizing, playing alphabet games, and learning about accents and syllabication. These activities not only reinforce the alphabet sequence, they also develop retrieval skills.

- **Review:** Each lesson includes some type of review following the warm-up activity. Typically these review letters, sounds, and spelling. These should be quick reviews that fit within the total 10 minute warm-up time.

√ **NEW INCREMENT (10 MINUTES):** This element of the program occurs after the Lesson Warm-up. Follow the outline in the lesson booklet in the Teacher's Manual. While the program recommends that teachers use as many of the examples as possible, teachers may pick and choose which examples they feel are most appropriate for their class. It is only mandatory that teachers do as many examples as are necessary for their students to demonstrate understanding.

- **Keywords and Deckcards, related Wall Cards** are introduced during this time to help reinforce the letters, their sounds, and written forms. When a new increment has a Deck Card/Letter Tile/Wall Card etc. that reinforces it, then that/those cards are introduced and added with the previous ones. However, the practice and review of prior letter/sounds will be during the Lesson Warm Up.
- **Student Spelling Dictionary and Reference Booklets** should also be used for students to practice looking up relevant information.

√ **APPLICATION AND CONTINUAL REVIEW (25 MINUTES FOR 1ST GRADE; 15 MINUTES FOR 2ND GRADE)** This section of the lesson consists of whole group board work, worksheets or other activities that may vary by lesson as well as assigning homework. This portion of the lesson may take even less time if there is not a decodable reader that goes with a particular lesson.

- board work **(approximately 5 minutes):** Before worksheets are distributed, the class as a whole codes and reads carefully chosen words, phrases and sentences on the board. The correctly coded examples should remain on the board for children to refer to when completing their worksheets.
- **Worksheet:** Four days a week, children will be given worksheets to complete in class. Worksheets are designed to be completed by children independently as a whole-class activity. The backside of the worksheet—the homework—may be completed in class if time permits or completed at home and returned the next day. If some children in your class need extra support or are greatly challenged consider working with them in small groups or individually.

  Because worksheets are designed to be a part of the initial learning experience and not a reflection of what children have mastered, worksheets should **never** be graded. However, be sure to check worksheets and have him/her correct any errors before sending it home.

√ **CLASSROOM PRACTICE (AT LEAST 15 MINUTES):** Classroom Practice and Fluency Practice are times for children to improve their weaknesses or apply their learning by playing **Kid Card** games, engaging in **independent reading**, or practicing their **decoding and fluency skills** with readers.

- **Games/Activities/Reading and other practice:** Each lesson will offer suggestions for games and other classroom activities appropriate for practice in selected areas. Teachers can select the activities they feel best fit the needs of their class and fit within the allotted and time.

√ **SCHOOL/HOME REINFORCEMENT:** *Teachers may choose whether or not to use this component of the program.* Provided at the end of each lesson is a list of materials that may be sent home at the end of the day. Use this list to ensure that children are taking home the necessary work to finish, read or review with their parents.

NOTE: The school/home reinforcement portion of the program is not a required and can be considered optional.

ACTIVITY TO BE ONCE PER WEEK:

√ **ASSESSMENT AND REMEDIATION:** Assessments are designed to be diagnostic tools rather than grading tools. Their purpose is to help you monitor the progress of each child, identify concepts that need more review, and decide what, if any, changes in the instruction pace should be made.

- **Oral and Written Assessments:** Phonics and spelling assessments occur after every fourth lesson and are both written and oral. Written sections are teacher-directed and given to the class as a group; oral portions are short, individual interviews that may take place at any time during the day or may be spread over several days. Because the oral assessments are administered to children independently, be sure to plan an activity for the other children during these interviews.
- **Site Word Evaluations:** Sight word evaluations are also built into the program to assess children's mastery of those words they must memorize. Each evaluation should be conducted with children individually; you might find it most efficient to complete them along with the oral assessments.

To reiterate, the written section occurs every fifth lesson; part of this assessment is oral. Teachers can take up to five days to administer this oral section with the class, as long as this section is completed prior to the next assessment. Teachers can also administer the site-word test along with the oral test since both are administered individually.

* For more information on assessments and evaluating assessments see page 35 of the Instructional Overview in your Teacher's Manual.

OPTIONAL PROGRAM COMPONENTS:

Naturally, you are encouraged to incorporate as many of the program components as possible. This will enhance your student's opportunity for learning and assist them in mastering important phonics and spelling concepts. However, there are parts of the program that can be considered optional and altered when time does not permit that you cover everything:

- **Handwriting Instruction**
- **Fluency and Decodable Readers:** These components can be considered optional. However, it is recommended that the fluency readers be used, unless there is a fluency portion in your basal reading program that is consistently covered. It is recommended that some fluency reading take place in your classroom.
- **Fluency Assessments:** The fluency assessments are used if the fluency piece is being implemented (this is an optional component of the program). Conducting these formal and informal reading fluency assessments will help you measure children's reading abilities, identify areas for reinforcements, and monitor their progress. You can use the results of these assessments to guide instruction as you help children work toward fluency benchmarks.
- **School/Home Reinforcement:** Again, teachers may choose whether or not to use this component of the program to provide additional practice and reinforcement for students at home. Teachers may use the list provided at the end of each lesson to see what materials that may be sent home at the end of the day. Use this list to ensure that children are taking home the necessary work to finish, read or review with their parents.

Appendix E

Technical Appendix

OVERVIEW OF THE TECHNICAL APPENDIX

The purpose of this appendix is to provide fellow researchers with additional technical information to fully evaluate the scientific rigor of this study. Specifically, this appendix is written for technical audiences so that they may examine the statistical procedures employed as well as make more informed judgments of the internal and statistical conclusion validity of this study. It is *not* written for lay people. This *Technical Appendix* contains the following information:

- Analytical goals of these analyses
- Analytical framework
- Results of data analyses by analytical framework

ANALYTICAL GOALS

The evaluation of the *Saxon Phonics and Spelling* program focuses on the following broadly-framed goals:

1. *Assessment of effectiveness of the Saxon Phonics and Spelling Program*
   Because the *Saxon Phonics and Spelling* is designed as a supplemental program, the effectiveness of this program is examined in comparison to classes that did not use *Saxon Phonics and Spelling*. The analytical framework used to identify the effectiveness of the *Saxon Phonics and Spelling* program is *causal* in a numbers of ways:

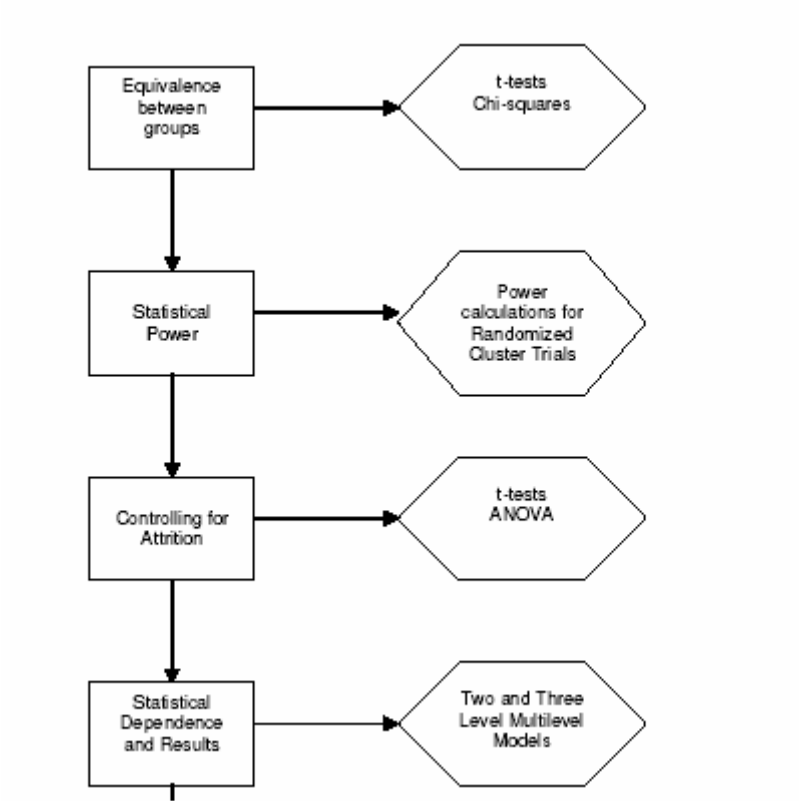
   (i) As described in the body of this final report, a well-planned randomized control trial was implemented;
   (ii) The analytical procedures pay close attention to multiple threats to internal validity including selection effects and attrition (Shadish, Cook, and Campbell, 2002);
   (iii) Given that students are "nested" within classrooms, the data are unlikely to be independent across students; dependence in outcomes is modeled by implementing hierarchical linear models (Raudenbush and Bryk, 2002);

2. *Knowledge development:* The implemented design also provides an opportunity to examine student and classroom/program measures that may be associated with program effectiveness for the *Saxon Phonics and Spelling* program. This relationship between student and teacher characteristics and program effectiveness is viewed as primarily associative and not causal for two reasons: (a) The implemented design is focused on estimating causal *main effects* for the program; the statistical power to identity program effects within subgroups is much lower; (b) There have been very few studies that have examined subgroup effects relating to curriculum of the *Saxon Phonics and Spelling* program as well as supplemental phonics and spelling interventions as a whole. In the absence of a strong program theory, the subgroup effects are viewed as empirical patterns that need theoretical frameworks and other rigorous experimental designs in the future to be estimated "causally."

ANALYTICAL FRAMEWORK

Figure E1 below and accompanying narrative show the five-step analytical procedures that were implemented to evaluate the effectiveness of the *Saxon Phonics and Spelling* program.

Figure E1. Description of Analytical Framework



(i) *Establishing group equivalence*: The differences in the treatment and control group were examined by conducting t-tests and chi-square analyses at the student and teacher levels on a range of baseline outcomes and other student and teacher characteristics. Care was taken to ensure that measures on which the groups differed significantly were used as covariates in subsequent analyses.

(ii) *Statistical power*: Dependency in the data decreases the statistical power to detect significant differences. Specifically, increased values of intra-class correlations (higher dependency in the data) results in reductions in statistical power. The power to detect significant differences in clustered random trials was calculated for a range of intra-class correlations and effect sizes, and also with and without a cluster covariate.[45]

(iii) *Controlling for attrition*: In this step, consideration is given to attrition as a potential threat to both internal and external validity of the study (Cook and Campbell, 1979).

[45]The use of a cluster-level covariate that is correlated with the outcomes of interest increases the power of the test (Raudenbush et al., 2005).

Both issues of measurement attrition (i.e., missing data due to student absences or lack of test administration) and dropout attrition (i.e., missing data due to students leaving the study) were examined.

Measurement Attrition

First, chi-square analysis was performed to determine if the proportion of measurement attrition was equivalent among both groups. In other words, this analysis examined whether there was a significant relationship between students who provided and did not provide data (at *each* time point) and group assignment (treatment vs. control). Second, ANOVAs were run to determine whether there were performance differences between those who completed the tests and those who did not by group using posttest measures (to examine those not providing pretest measures) and pretest measures (to examine those not providing posttest measures). An interaction between group and test completion status would be indicative of a bias because the type of treatment students who did not complete the test would be different than the type of control students who did not complete the test.

Dropout Attrition

The potential problems of overall attrition and differential attrition due to students leaving the study was first "diagnosed" using a simple statistical procedure; specifically, chi-square analysis was conducted to determine if the proportion of dropout attrition was equivalent among both groups. Second, in order to determine whether there was differential attrition on pretest measures, ANOVAs were run to determine if there was (1) a significant interaction between group and attrition status, and (2) a significant main effect for attrition status (Cook and Campbell, 1979). A significant interaction would indicate a threat to internal validity because the type of student dropping out of the treatment group would be different than the type of student dropping out of the control group. A significant main effect would indicate a threat to external validity because the students remaining in the study would be different than the students who dropped out of the study.

(iv) *Statistical Dependency and Results:* Three-level multilevel models (Raudenbush and Bryk, 2002) were first implemented to examine growth in scores between the pre and post periods in the treatment group only. Since the purpose of this particular analysis was not to explain changes in performance, this initial set of growth models for the treatment group did not include any covariates. The mathematical details of this model are presented in Appendix F.

Next, three-level multilevel models were implemented to estimate program effects. Both student-level and teacher-level covariates were included in the multilevel models. In addition, key teacher and student characteristics were identified based on prior educational research as related to educational outcomes, and included in the models. In the three-level model, student outcomes were modeled at level 1, student characteristics were modeled at level 2, and teacher/school characteristics were modeled at level 3. Appendix F describes the mathematical equations representing the three-level multilevel models.

RESULTS

This section is organized according to the aforementioned analytical framework.

1. ESTABLISHING GROUP EQUIVALENCE

a) The relationship between various student demographic variables and group status was examined. Results showed that there were significantly more 2nd graders and less 1st graders in the *Saxon* group as compared to control group. In addition, there were a higher proportion of control students classified as being high-performing in literacy as compared to *Saxon* students. More detailed information on these analyses is presented in Table 3 in the main body of this report.

b) Pre-test differences on the assessment measures were examined. Data for 1st and 2nd graders were combined when possible as noted. Student level analyses showed that there were no statistically significant differences, at the $p<.05$ level, between the groups out of the 6 pretest scores (see Table E1).

Table E1. Sample Size, Means, Standard Deviations, and *t*-test (Student Level) Results for Assessments at Pre-testing

Pretest*	Group	N	Mean	Std. Dev.	*t*	Sig. Level
ITBS – Word Analysis SS (1st and 2nd graders)	Control	304	149.33	20.84	0.859	.39
	Treatment	363	147.99	19.26		
ITBS – Reading Words SS (1st grade only)	Control	164	139.43	11.38	-0.283	.78
	Treatment	164	139.82	13.88		
ITBS - Spelling SS (2nd grade only)	Control	140	156.95	12.73	1.646	.10
	Treatment	199	154.83	10.91		
Morris & Perney Spelling Test – PCT (1st grade only)	Control	151	58.57	23.08	-0.530	.60
	Treatment	163	59.98	24.15		
Ganske Spelling Test – PCT (1st and 2nd graders)	Control	315	16.52	13.95	-1.218	.22
	Treatment	339	17.82	13.20		
Dolch Words – PCT (1st and 2nd graders)	Control	302	63.17	36.43	-1.873	.06
	Treatment	355	68.30	33.28		

*SS=Scale Score; PCT=Percent of total

c) Analyses were also performed to examine whether baseline differences existed as a function of teacher characteristics. Results showed no significant baseline differences among teachers in terms of degree earned, $\chi 2=1.427$, $p=.49$, or overall teaching experience, $t(33)=.10$, $p=.93$, and at their current grade in particular, $t(33)=1.039$, $p=.31$. There were also no differences on affective measures such as perceptions of control over instructional

decisions, $t(33)=.51, p=.61$, and barriers to effective teaching $t(33)=1.747, p=.09$. However, significant differences were observed in terms of their knowledge and preparation to teach via the five elements of reading, and spelling and writing, $t(32)=4.087, p<..001$, and their engagement in effective literacy practices, $t(33)=3.459, p<.001$. In particular, results showed that at baseline, control teachers reported having more knowledge and preparation for teaching important elements of reading (fluency, phonics, phonemic awareness, comprehension and vocabulary), spelling, and writing, and engaged in effective literacy practices to a greater extent than treatment teachers. Given these significant differences, analyses comparing treatment and control groups controlled for these two variables.

d) Classroom environment was also analyzed based on information collected during the Fall. Again, results showed no significant differences between treatment and control teachers in terms of how their classes were structured (whole group v.s. small groups v.s. individuals), $p>.05$ for all items, availability of instructional materials, $p>.05$ for all items, design of their lessons, $p>.05$ for all items, teacher-student interactions, $t(30)=.152, p=.88$, class culture, $t(30)=.485, p=.63$, general lesson content, $t(30)=.324, p=.75$, class engagement, $t(30)=1.086, p=.29$, and overall class climate, $t(30)=.272, p=.79$. In terms of their core reading and language arts time, results showed no significant differences in the amount of time spent on reading, writing, spelling, and phonics activities, $p>.05$ for all items. However, with the inclusion of time associated with *Saxon Phonics and Spelling* instruction, results did reveal a significant difference, with treatment teachers devoting more time (24 additional minutes) in reading and language arts instruction (including phonics and spelling) than control teachers, $t(33)=2.72, p=.01$.

2. STATISTICAL POWER

The following assumptions were used to calculate the power to detect effects:

- Significance level (α) = 0.05;
- 35 clusters (teachers) with an average class size of 21.
- Calculations were done both without and with a cluster covariate. A cluster-level covariate that was correlated with individual-level outcome was available; specifically, the average class-level score at baseline was correlated with the ITBS ($r_{wordanalysis}$= 0.53, $r_{readwords}$=.20, $r_{spelling}$=.22), spelling test ($r_{m\&p}$=.84 and r_{ganske}=.43) and Dolch total scores (r = 0.28). The average value of 0.42 was taken in these calculations.
- The calculations were done on a range of intra-class correlations. The unadjusted intra-class correlation for the ITBS, spelling, and Dolch Word tests at baseline ranged from .11 to .46 respectively.

The *Optimal Design* software was used in the calculations in this section (Raudenbush et al., 2005). This program is designed to determine the power of longitudinal and multilevel research. Figure E2 describes the power for a cluster randomized trial for a range of intra-class correlations *without* any cluster covariate for low, medium and high power (effect sizes corresponding to 0.2, 0.5, and 0.8 respectively). Figure E3 describes the power for a cluster randomized trial with a correlated cluster variable (r = 0.42). The key point from the graphics

below is that there is enough power *with* a cluster covariate to reasonably detect a medium to large effect size.

Given the similarities in reading content and practices employed by control and treatment teachers and the year-long duration of the study, small to moderate effect sizes were expected. It should also be noted that prior educational research studies show that research in these applied settings tends to have more "noise" in terms of student outcomes being related to numerous teacher, class, and school factors that cannot all be measured. Furthermore, according to Slavin (1986), a leader in educational evaluation, an effect size of 0.25 is considered to be educationally significant.

Figure E2. Power vs. Intra-Class Correlations for a Range of Effect Sizes (No Cluster-Level Covariate Included)

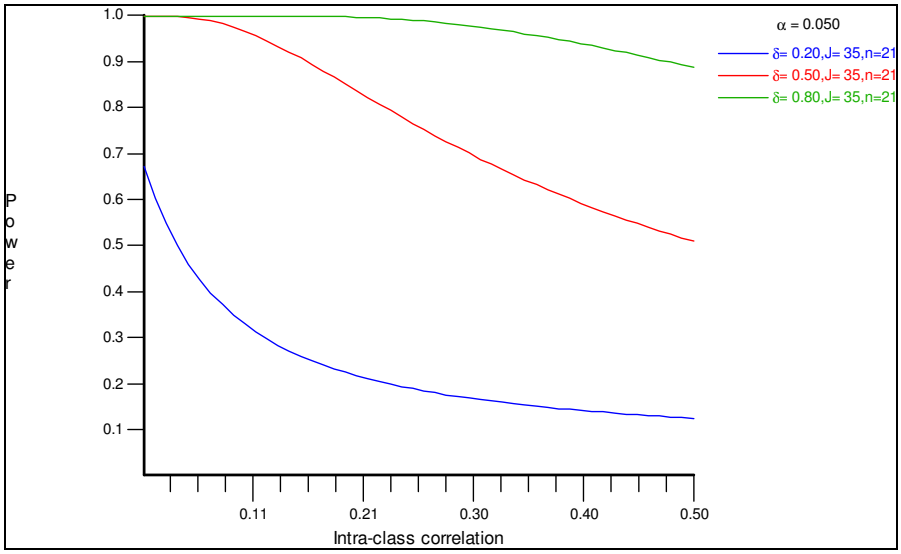
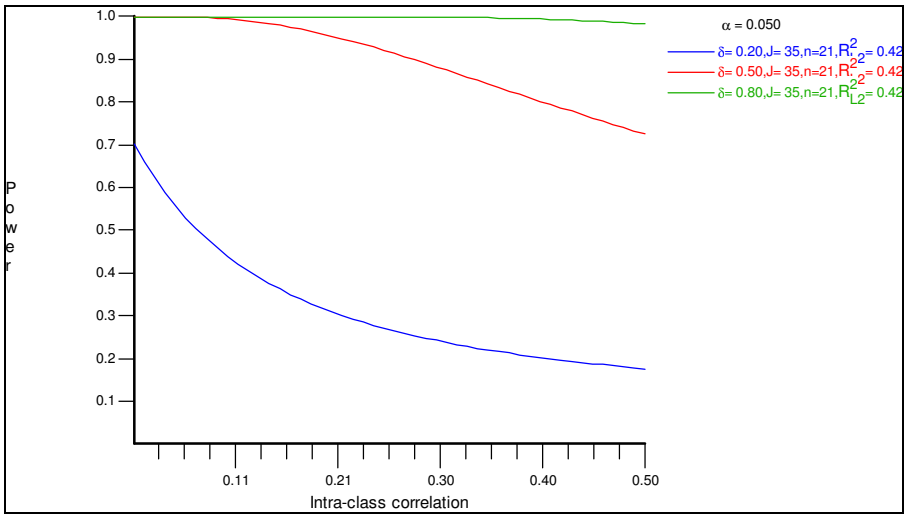


Figure E3. Power vs. Intra-Class Correlations for a Range of Effect Sizes (Cluster-Level Covariate Included)



Note: In Figures E2 and E3, J refers to number of clusters, n refers to the average cluster size, δ refers to the effect size, α is the significance level, and r_2 is the correlation coefficient between the cluster-level covariate and the individual-level outcomes.

3. ATTRITION ANALYSIS

As previously noted, both measurement attrition (i.e., missing data due to students not completing assessments) and dropout attrition (i.e., missing data due to students leaving the study) were examined. The approach taken in this project was to seek a consistent pattern of results of program effects across a range of methods. In this section, the observed pattern of differential attrition is examined to determine if it can explain the pattern of the observed results.

Measurement Attrition

A portion of the students did not have data available at pre or post test due to absences on test administration days or because the test was simply not administered. This occurred despite multiple contacts made to teachers by researchers to have all students administered. A second grade control teacher at school F did not administer the Dolch Word test during post-testing. This teacher did administer all other post-tests, however[46]. Table E2 lists the number (and percent) of students who were in the study throughout the school year but did not provide pre or post tests. For the most part, teachers did very well in ensuring that all subtests were administered.

Chi-square analyses showed one significant relationship. Specifically, there were more control students who did not take the Dolch Word test as compared to *Saxon Phonics and Spelling* students. As previously mentioned, this is due to the fact that one control teacher did not administer the Dolch test to the class. There were no other significant relationships between the proportion of students who provided and did not provide data and group at both pre and post test. Taking into consideration all measures, these results indicate that the *type* of student (control versus treatment) not taking the tests at pre and post testing was not substantively different.

Furthermore, to examine if there were any *performance* differences between those who completed tests and those that did not by group, ANOVAs were run on the post-test measures (to examine those not providing pretest measures) and on pretest measures (to examine those not providing posttest measures). Significant interactions between measurement attrition status and group assignment would suggest a bias. Results showed no significant interactions (see Table E2) *out of the 12* comparisons. This suggests that measurement attrition is unlikely to bias results.

[46] This was discovered when data was received by PRES Associates for data entry. By this point, the school year had ended and retesting could not be done.

Table E2. Number of Students Who Did Not Provide Pre and Post Data

	Admin Time	N (%) Who Did Not Take Test			Chi-Square	ANOVA for interaction
		Treatment	Control	Total		
ITBS* (1st and 2nd graders)	Pre (N=667)	7 (1.9%)	8 (2.6%)	15 (2.2%)	$\chi2(1)=0.36$, p=0.55	$F_{wordanalysis}$ (1, 644)=0.119, p=0.730 $F_{readingwords}$ (1, 311)=0.153, p=0.696 $F_{spelling}$ (1, 332)=3.512, p=0.062
	Post (N=654)	18 (4.9%)	10 (3.2%)	28 (4.1%)	$\chi2(1)=1.18$, p=0.28	$F_{wordanalysis}$ (1, 663)=0.072, p=0.789 $F_{readingwords}$ (1, 324)=1.146, p=0.285 $F_{spelling}$ (1, 335)=2.562, p=0.110
Spelling Tests* (1st and 2nd graders)	Pre (N=661)	9 (2.4%)	12 (3.8%)	21 (3.1%)	$\chi2(1)=1.13$, p=0.29	$F_{M\&P\ Spelling}$(1, 319)=1.518, p=0.219 $F_{Ganske\ spelling}$ (1,654)=0.530, p=0.467
	Post (N=639)	19 (5.1%)	24 (7.7%)	43 (6.3%)	$\chi2(1)=1.87$, p=0.17	$F_{M\&P\ Spelling}$(1, 310)=2.305, p=0.130 $F_{Ganske\ spelling}$ (1,650)=0.021, p=0.886
Dolch Words (1st and 2nd graders)	Pre (N=665)	8 (2.2%)	9 (2.9%)	17 (2.5%)	$\chi2(1)=0.36$, p=0.55	F(1, 646)=1.217, p=0.270
	Post (N=631)	15 (4.1%)	36 (11.5%)	51 (7.5%)	$\chi2(1)=13.70$, p<0.001	F(1, 679)=0.153, p=0.696

*Students not providing a pre or post ITBS or spelling test did so across all subtests. Therefore, one attrition measure is presented for ITBS and the Spelling tests.

Dropout Attrition

There was an overall attrition of 9% due to students leaving school or shifting from control to treatment classes (or vice-versa). Analyses were performed to examine if there was *differential attrition* as a result of students leaving. First, analyses were performed to examine if the proportion of dropout attrition was equivalent among both groups. As shown in Table E3, results showed that this was the case.

Table E3. Number of Students by Enrollment Status*

	Students		
	Control	Treatment	Total
Total participants on roster	342 (100.0%)	407 (100.0%)	749 (100%)
Students who moved/left	30 (8.8%)	37 (9.1%)	67 (8.9%)
Total students remaining throughout school year	312 (91.2%)	370 (90.9%)	682 (91.1%)

*$\chi 2$ (1)= 0.023, p =.88

Secondly, analyses were performed to examine whether baseline performance differences existed between students who remained in the study and those who left and group assignment. Of interest in these ANOVAs were the interactions of group assignment and attrition status and the main effect for attrition status. A significant interaction would indicate a threat to internal validity. Similarly, a main effect for attrition status would suggest a threat to external validity.

Examination of the *interactions* showed no significant group by attrition status interactions on all measures. Analysis of the *main effects* for attrition status showed three main effects as measured by the ITBS Word Analysis, Ganske Spelling, and Dolch Word tests, see Table E4. Results showed that attrition was associated with lower scores on these tests. However, the small number of students who left or moved are also of note.

Table E4. ANOVA Results for Pre-Tests by Group and Attrition Status

Measure	Attrition Status	Group	N	Mean	Sd.	ANOVA for interaction	ANOVA for main effect
ITBS – Word Analysis SS (1st and 2nd graders)	Attrition	Control	13	138.15	14.01	F(1, 692)=0.003, p=0.956	F(1, 692)=8.92, p=.003
		Treatment	16	136.39	23.39		
	No change	Control	304	149.33	20.84		
		Treatment	363	147.99	19.26		
ITBS – Reading Words SS (1st grade only)	Attrition	Control	9	136.56	5.85	F(1, 342)=0.051, p=0.821	F(1, 342)=1.340, p=0.248
		Treatment	9	135.56	18.35		
	No change	Control	164	139.43	11.38		
		Treatment	164	139.82	13.88		
ITBS - Spelling SS (2nd grade only)	Attrition	Control	4	154.75	4.50	F(1, 346)=0.196, p=0.658	F(1, 346)=1.075, p=0.300
		Treatment	7	149.35	10.56		
	No change	Control	140	156.95	12.73		
		Treatment	199	154.83	10.91		
Morris & Perney Spelling Test – PCT (1st grade only)	Attrition	Control	7	57.86	8.64	F(1, 326)=1.041, p=0.308	F(1, 326)=1.292, p=0.257
		Treatment	9	46.85	28.06		
	No change	Control	151	58.57	23.08		
		Treatment	163	59.98	24.15		
Ganske Spelling Test – PCT (1st and 2nd graders)	Attrition	Control	9	6.11	6.97	F(1, 675)=0.206, p=0.650	F(1, 675)=10.164, p=0.001
		Treatment	16	10.00	12.78		
	No change	Control	315	16.52	13.95		
		Treatment	339	17.82	13.20		
Dolch Words – PCT (1st and 2nd graders)	Attrition	Control	12	37.40	27.85	F(1, 679)=0.122, p=0.727	F(1, 679)=11.230, p=0.001
		Treatment	14	47.40	37.86		
	No change	Control	302	63.17	36.43		
		Treatment	355	68.30	33.28		

In summary, there was very limited evidence for dropout attrition and measurement attrition. While dropout attrition was associated with lower scores on three tests, this was consistent across both groups (differential attrition was not observed). Furthermore, the lack of Dolch post-testing by one teacher resulted in less control students taking the Dolch Word test as compared to *Saxon Phonics* students. However, there were no *performance* differences between group and those who completed tests and those that did not. Overall, results suggest that attrition is unlikely to bias results.

4. Statistical Dependency and Results

Multilevel models were run to assess growth in outcomes over time in the treatment group and also to assess program effects. However, prior to running these analyses paired sample t-tests were run to obtain preliminary information on changes in performance from pre-testing to post-testing.

Analysis of Growth among Treatment Students

t-tests for Change from Pretest to Posttest

Table E5 presents the means obtained for treatment students using the *Saxon Phonics and Spelling* program at pre- and posttest as measured by the ITBS, M&P and Ganske Spelling tests, and Dolch Word assessments. Paired sample t-tests were conducted to examine whether there was significant change from pretest to posttest. Results showed significant growth (i.e., improvement in performance) on all outcome measures. However, this analysis is only intended to be descriptive and as described subsequently other explanations for such improvement in test scores (e.g., the core reading and language arts curricula) also need to be considered.

Table E5. Pre-Post Scores for Treatment Students (Paired Sample t-test Results)

	N	Pre Mean (sd)	Post Mean (sd)	t	Sig. Level
ITBS – Word Analysis SS (1st and 2nd graders)	343	148.52 (19.36)	169.53 (25.19)	19.921	<.001
ITBS – Reading Words SS (1st grade only)	152	140.22 (14.00)	153.53 (11.48)	12.937	<.001
ITBS - Spelling SS (2nd grade only)	192	155.15 (10.84)	174.28 (13.81)	20.161	<.001
Morris & Perney Spelling Test – PCT (1st grade only)	149	61.25 (23.32)	79.13 (18.11)	10.011	<.001
Ganske Spelling Test – PCT (1st and 2nd graders)	337	17.97 (13.03)	66.05 (23.69)	48.923	<.001
Dolch Words – PCT (1st and 2nd graders)	340	68.24 (33.57)	92.42 (15.16)	15.482	<.001

Furthermore, these analyses do not take into account the hierarchical nature of these data nor do they examine the treatment group relative to the control group—both critical in making causal claims. Therefore, multilevel model analyses were performed to model the growth of treatment students nested within classrooms. The next section addresses the hierarchical nature of the data.

Multilevel Models for Growth in Treatment Students

Because the purpose of this set of analyses is not explanatory (i.e., to explain why there was growth—this is dealt with in the following section) the model does not include any covariates. Of interest in the two-level model is the slope from pre- to post-test at level 1. Level 2 includes only the random effect associated with the teacher nesting factor. Appendix F describes the mathematical details of this model.

The multilevel model was run on each of the outcomes. Table E6 summarizes the results of these analyses. Note that each row in Table E6 corresponds to the growth coefficients estimated for that dependent variable from a *separate* multilevel model. Similar to that found in Table E5, there was significant improvement in scores on the all outcomes.

Table E6. Treatment Group Growth from Two-Level Models*

Predictor		Coefficient	Standard Error	t-ratio	Sig. Level
ITBS – Word Analysis SS (1st and 2nd graders)	Level 2 Slope-β10	21.20	1.33	15.970	<.001
ITBS – Reading Words SS (1st grade only)	Level 2 Slope-β10	13.58	1.37	9.924	<.001
ITBS - Spelling SS (2nd grade only)	Level 2 Slope-β10	19.55	1.13	17.258	<.001
Morris & Perney Spelling Test – PCT (1st grade only)	Level 2 Slope-β10	17.83	2.13	8.364	<.001
Ganske Spelling Test – PCT (1st and 2nd graders)	Level 2 Slope-β10	47.79	1.17	40.787	<.001
Dolch Words – PCT (1st and 2nd graders)	Level 2 Slope-β10	24.13	1.70	14.158	<.001

*Each row corresponds to the coefficients estimated from running a *separate* model for each of the total tests and subtests.

Growth Analysis of Subgroups of Treatment Students

Exploratory analysis was also performed to examine the relationship between the *Saxon Phonics and Spelling* program and subgroup performance. That is, the results summarized in this section deal with the performance among treatment students only. It is important to note that due to the small sample sizes and likely dependency issues, no causal, conclusive statements should be made. Nevertheless, these results are presented for preliminary, exploratory purposes. In addition, only statistically significant results (at the .03 level) are presented with accompanying statistics[47]. Results showed significant interactions for the following subgroup categories: ethnicity, grade level, free/reduced lunch status, and ability level. There were no significant interactions observed for gender, English language status, and special education status.

[47] These significance levels have been adjusted for multiple comparisons using the procedure outlined by Sankoh and colleagues (1997).

The accompanying tables (E7-E14) include the repeated measures ANOVA analyses whereby the interaction of group and time is of interest. A significant finding indicates that change in test scores over time depends on whether students were in the subgroups or not. In addition, paired t-tests' results are also presented. For these latter analyses, only treatment students within these subgroups are included. This provides preliminary information on whether students in these subgroups show growth in phonics, spelling, and reading performance.

Ethnicity

Repeated measures ANOVA results showed a significant interaction between time (pre to post) and ethnicity (White versus Hispanic versus African American), see Table E7. Specifically, White *Saxon Phonics & Spelling* students showed the greatest change, followed by Hispanics and African Americans, as measured by the ITBS Word Analysis, ITBS Spelling, M&P Spelling, and Ganske Spelling scores. There were no significant interactions on the remaining two assessment scores. Furthermore, exploratory paired sample t-tests showed significant growth among Whites, Hispanics and African Americans on all assessments, see Table E8.

Table E7. Repeated Measures ANOVA for Treatment Students by Ethnicity

	F*	df	Sig. value
ITBS – Word Analysis (1^{st} and 2^{nd} graders)	10.434	2, 330	<.001
ITBS – Spelling (2^{nd} graders)	13.559	2, 185	<.001
Morris & Perney Spelling Test (1^{st} graders)	5.706	2, 139	.004
Ganske Spelling Test (1^{st} and 2^{nd} graders)	14.098	2, 324	<.001

*F for interaction of time and group.

Table E8. Paired t-test Results for Treatment Students by Ethnicity

	N	Pre Mean (sd)	Post Mean (sd)	*t*	Sig. Level
White Students' Results					
ITBS – Word Analysis SS (1st and 2nd graders)	145	152.16 (21.27)	178.77 (26.59)	16.055	<.001
ITBS – Reading Words SS (1st grade only)	54	140.74 (15.84)	155.44 (12.14)	8.158	<.001
ITBS - Spelling SS (2nd grade only)	91	155.24 (10.66)	179.27 (11.80)	19.534	<.001
Morris & Perney Spelling Test – PCT (1st grade only)	53	59.21 (25.30)	80.72 (15.93)	7.818	<.001
Ganske Spelling Test – PCT (1st and 2nd graders)	144	20.97 (13.17)	74.62 (19.75)	45.631	<.001
Dolch Words – PCT (1st and 2nd graders)	148	72.20 (33.59)	94.50 (13.55)	9.319	<.001
Hispanic Students' Results					
ITBS – Word Analysis SS (1st and 2nd graders)	93	143.09 (17.89)	161.24 (21.51)	10.591	<.001
ITBS – Reading Words SS (1st grade only)	45	138.64 (1.95)	152.87 (10.96)	8.731	<.001
ITBS - Spelling SS (2nd grade only)	49	151.48 (11.95)	166.78 (13.46)	8.734	<.001
Morris & Perney Spelling Test – PCT (1st grade only)	43	62.33 (22.26)	82.48 (17.32)	6.763	<.001
Ganske Spelling Test – PCT (1st and 2nd graders)	90	14.3 (11.54)	60.06 (22.65)	22.398	<.001
Dolch Words – PCT (1st and 2nd graders)	92	60.20 (34.47)	89.37 (18.30)	10.136	<.001

Table E8 Continued

African American Students' Results					
ITBS – Word Analysis SS (1st and 2nd graders)	95	148.67 (16.18)	164.67 (21.87)	7.671	<.001
ITBS – Reading Words SS (1st grade only)	47	141.89 (13.35)	151.28 (11.35)	4.952	<.001
ITBS - Spelling SS (2nd grade only)	48	158.58 (9.14)	172.63 (13.82)	7.163	<.001
Morris & Perney Spelling Test – PCT (1st grade only)	46	66.09 (18.91)	74.86 (20.85)	2.864	.006
Ganske Spelling Test – PCT (1st and 2nd graders)	93	17.69 (13.05)	59.52 (26.66)	20.131	<.001
Dolch Words – PCT (1st and 2nd graders)	91	72.13 (30.25)	91.78 (14.35)	6.891	<.001

Grade Level

The ITBS Word Analysis, Ganske Spelling, and Dolch Word List tests were administered to both 1st and 2nd graders. The remaining assessments were administered to 1st *or* 2nd graders. Thus, researchers were able to determine if differences existed between the 1st and 2nd graders on the former tests only. Repeated measures ANOVA results showed a significant interaction between time (pre to post) and grade level, see Table E9. Specifically, 2nd grade *Saxon Phonics & Spelling* students showed greater change on the ITBS Word Analysis and Ganske Spelling tests as compared to 1st graders. In contrast, 1st graders showed greater change than 2nd graders on the Dolch Word List test. However, this may be due to the high percentage of correct responses observed among 2nd graders at pretesting (Dolch words were noted as too easy) and therefore, there was less room for growth. Furthermore, exploratory paired sample t-tests showed significant growth among both 1st and 2nd graders on all assessments, see Table E10.

Table E9. Repeated Measures ANOVA for Treatment Students by Grade Level

	F*	df	Sig. value
ITBS – Word Analysis (1st and 2nd graders)	7.273	1, 341	.007
Ganske Spelling Test (1st and 2nd graders)	6.948	1, 335	.009
Dolch Word Test (1st and 2nd graders)	114.008	1, 338	<.001

*F for interaction of time and group.

Table E10. Paired t-test Results for Treatment Students by Grade Level

	N	Pre Mean (sd)	Post Mean (sd)	*t*	Sig. Level
1st Graders' Results					
ITBS – Word Analysis SS (1st and 2nd graders)	152	138.12 (16.76)	155.97 (18.70)	12.921	<.001
Ganske Spelling Test – PCT (1st and 2nd graders)	149	10.17 (10.60)	55.37 (22.38)	30.050	<.001
Dolch Words – PCT (1st and 2nd graders)	147	48.68 (33.52)	89.40 (16.36)	16.379	<.001
2nd Graders' Results					
ITBS – Word Analysis SS (1st and 2nd graders)	191	156.79 (17.20	180.32 (24.50)	15.467	<.001
Ganske Spelling Test – PCT (1st and 2nd graders)	188	24.15 (11.37)	74.52 (21.18)	39.465	<.001
Dolch Words – PCT (1st and 2nd graders)	193	83.14 (24.88)	94.71 (13.79)	8.002	<.001

Free/Reduced Lunch Status

Repeated measures ANOVA results showed a significant interaction between time (pre to post) and free/reduced lunch status, see Table E11. Specifically, *Saxon Phonics and Spelling* students not receiving free/reduced lunch status showed greater change as measured by the ITBS Word Analysis, ITBS Spelling, and Ganske Spelling tests as compared to students receiving free/reduced lunch. However, the reverse relationship was observed for the Dolch Word List test. That is, students receiving free/reduced lunch showed greater change on high frequency word reading than students not receiving free/reduced lunch. Exploratory paired sample t-tests also showed significant growth among both types of students on all assessments, see Table E12.

Table E11. Repeated Measures ANOVA for Treatment Students by Free/Reduced Lunch Status

	F*	df	Sig. value
ITBS – Word Analysis (1st and 2nd graders)	8.864	1, 225	.003
ITBS – Spelling (2nd graders)	5.385	1, 128	.022
Ganske Spelling Test (1st and 2nd graders)	19.122	1, 221	<..001
Dolch Word Test (1st and 2nd graders)	13.543	1, 221	<..001

*F for interaction of time and group.

Table E12. Paired t-test Results for Treatment Students by Free/Reduced Lunch Status

	N	Pre Mean (sd)	Post Mean (sd)	*t*	Sig. Level
Students Not Receiving Free/Reduced Lunch Results					
ITBS – Word Analysis SS (1st and 2nd graders)	96	158.23 (19.99)	184.45 (26.23)	12.402	<.001
ITBS – Reading Words SS (1st grade only)	32	144.88 (16.21)	160.44 (12.24)	6.480	<.001
ITBS - Spelling SS (2nd grade only)	64	157.83 (10.96)	180.69 (11.97)	13.185	<.001
Morris & Perney Spelling Test – PCT (1st grade only)	29	70.23 (19.27)	85.06 (14.63)	9.470	<.001
Ganske Spelling Test – PCT (1st and 2nd graders)	93	24.57 (11.97)	79.94 (18.54)	37.956	<.001
Dolch Words – PCT (1st and 2nd graders)	94	84.77 (24.69)	96.49 (10.20)	5.493	<.001
Students Receiving Free/Reduced Lunch Results					
ITBS – Word Analysis SS (1st and 2nd graders)	131	146.74 (16.52)	164.76 (22.76)	10.143	<.001
ITBS – Reading Words SS (1st grade only)	64	141.25 (13.72)	152.42 (11.26)	7.668	<.001
ITBS - Spelling SS (2nd grade only)	66	156.13 (8.53)	173.60 (12.22)	11.344	<.001
Morris & Perney Spelling Test – PCT (1st grade only)	64	65.34 (16.09)	74.61 (21.06)	4.874	<.001
Ganske Spelling Test – PCT (1st and 2nd graders)	130	17.69 (13.14)	62.69 (24.69)	26.259	<.001
Dolch Words – PCT (1st and 2nd graders)	129	67.80 (30.53)	91.78 (15.23)	10.084	<.001

Students at Various Literacy Levels

Literacy performance level for each student was determined by using pretest ITBS percentile rankings. The bottom 30% were categorized as low, 31-69% were categorized as moderate, and the top 70% were categorized as high[48]. Comparisons were made between the three identified literacy levels. Results showed a significant interaction between time and literacy level as measured by the ITBS Word Analysis, ITBS Reading Words, Morris & Perney Spelling test, and

[48] This measure was used because it was the only standardized, norm-referenced test that could provide information on performance levels relative to a national sample.

Dolch Word test, see Table E13. Specifically, the change in performance from pre to post was greatest among low-level students, followed by average and then high level students. Note that this may be due to the fact that there is more room for improvement among low-level students as compared to higher level students. Exploratory simple effects paired sample t-tests showed significant gains on all assessment measures, with the exception of high-level students on the ITBS Reading Words subtest, see Table E14.

Table E13. Repeated Measures ANOVA for Treatment Students by Literacy Skill Level

	F*	df	Sig. value
ITBS – Word Analysis (1^{st} and 2^{nd} graders)	5.864	2, 340	.003
ITBS – Reading Words (1^{st} graders)	19.366	2, 149	<.001
Morris & Perney Spelling Test (1^{st} graders)	7.487	2, 145	.001
Dolch Word Test (1^{st} and 2^{nd} graders)	42.573	2, 335	<.001

*F for interaction of time and group.

Table E14. Paired t-test Results for Treatment Students by Literacy Skill Level

	N	Pre Mean (sd)	Post Mean (sd)	*t*	Sig. Level*
Low Performing Students' Results					
ITBS – Word Analysis SS (1^{st} and 2^{nd} graders)	66	129.06 (10.82)	154.40 (21.30)	10.146	<.001
ITBS – Reading Words SS (1^{st} grade only)	29	125.62 (9.19)	146.21 (8.91)	8.781	<.001
ITBS - Spelling SS (2^{nd} grade only)	37	142.43 (7.98)	163.08 (13.59)	8.966	<.001
Morris & Perney Spelling Test – PCT (1^{st} grade only)	27	36.73 (23.86)	67.96 (19.93)	5.184	<.001
Ganske Spelling Test – PCT (1^{st} and 2^{nd} graders)	62	7.10 (6.93)	53.06 (22.26)	18.138	<.001
Dolch Words – PCT (1^{st} and 2^{nd} graders)	68	34.36 (32.83)	82.00 (24.12)	12.708	<.001

Table E14 Continued

Average Performing Students' Results					
ITBS – Word Analysis SS (1st and 2nd graders)	228	149.15 (15.83)	170.61 (23.94)	17.320	<.001
ITBS – Reading Words SS (1st grade only)	96	138.82 (8.07)	153.07 (10.77)	13.006	<.001
ITBS - Spelling SS (2nd grade only)	133	156.40 (8.58)	175.52 (12.30)	16.492	<.001
Morris & Perney Spelling Test – PCT (1st grade only)	96	63.42 (16.91)	78.11 (16.85)	7.941	<.001
Ganske Spelling Test – PCT (1st and 2nd graders)	225	18.38 (11.91)	66.16 (22.90)	40.060	<.001
Dolch Words – PCT (1st and 2nd graders)	222	73.55 (28.92)	94.17 (11.37)	11.846	<.001
High Performing Students' Results					
ITBS – Word Analysis SS (1st and 2nd graders)	49	171.78 (15.85)	184.88 (25.21)	4.499	<.001
ITBS – Reading Words SS (1st grade only)	27	160.89 (10.97)	163.00 (10.13)	.902	.376
ITBS - Spelling SS (2nd grade only)	22	169.00 (.000)	185.64 (10.07)	7.746	<.001
Morris & Perney Spelling Test – PCT (1st grade only)	25	81.87 (17.62)	95.27 (7.16)	5.152	<.001
Ganske Spelling Test – PCT (1st and 2nd graders)	48	30.32 (12.44)	82.60 (18.65)	22.535	<.001
Dolch Words – PCT (1st and 2nd graders)	48	92.60 (12.16)	98.82 (2.37)	3.649	.001

Change by *Saxon Phonics and Spelling* Implementation Levels

In order to provide preliminary information on the relationship between overall implementation of the *Saxon Phonics and Spelling* program (low, moderate, and high) and student performance, multilevel analyses were performed. It should be noted that these are exploratory and non-causal. Table E15 displays the results of these analyses. As shown, there was no relationship between level of implementation and growth in performance from pre- to post-testing. That is, students whose teachers used the *Saxon Phonics and Spelling* program improved on these measures, regardless of their level of implementation.

Table E15. Differences by Elements of Reading Implementation Level

	ITBS Word Analysis	ITBS Spelling	ITBS Reading Words	Ganske Spelling	M&P Spelling	Dolch Word List
Coefficient (Std. Err.)	3.51 (2.77)	1.29 (2.05)	-.18 (2.36)	-3.00 (2.55)	6.13 (3.98)	.84 (3.34)
t-ratio	1.268	.630	-.078	-1.177	1.539	.252
Sig	.21	.53	.94	.24	.13	.80

In order to provide preliminary information on the components of the *Saxon Phonics and Spelling* program and other teacher characteristics that are associated with greater gains in student performance (on ITBS), regression analyses were performed. To reiterate, these are exploratory and are only provided so that readers can have preliminary information on the *Saxon Phonics and Spelling* components that are related to the greatest levels of change. Table E16 displays the results of the stepwise regression model run. As shown, five variables significantly contributed to the final regression model, $F(5, 302)=19.386, p<.001$. The final model accounted for 23% of the variance in student performance gains. Note that with the exception of the extent to which teachers engaged in effective literacy practices, all variables included in the final model were *Saxon Phonics and Spelling* program components.

Table E16. Regression Results Between *Saxon Phonics and Spelling* Components and Gain in Student Performance

Component	Beta (Stnd. Coeff.)	Sig.
Included Variables:		
Percent of *Saxon* Phonics & Spelling lessons completed	0.378	<0.001
Assigned worksheet for students to complete	0.226	<0.001
Literacy Instructional Practices*	0.297	<0.001
Used student spelling dictionary & reference booklets	0.196	0.002
Used the fluency readers	0.135	0.026
Excluded Variables:	Beta In	Sig.
Amount of time spent on *Saxon* Phonics & Spelling lessons	0.069	0.389
Complete board work	-0.157	0.141
Conduct classroom practice (Kid Card, independent reading, decoding/fluency activities)	-0.017	0.755
Used handwriting activities	-0.100	0.241
Used fluency assessment	-0.057	0.461
Used decodable readers	-0.013	0.816
Introduced lesson	0.031	0.614
Engaged in language alpha activity	0.026	0.705
Reviewed letters, sounds and/or spelling	-0.015	0.804
Taught new increment	-0.066	0.777
Teacher Knowledge/Preparation for Literacy Instruction*	0.052	0.382
Years teaching*	0.042	0.461
Average amount of time spent on reading (outside of *Saxon*)	-0.003	0.967
Average amount of time spent on phonics (outside of *Saxon*)	-0.009	0.897
Average amount of time spent on spelling (outside of *Saxon*)	-0.087	0.141

*These variables are not directly related to the program itself.

Prior Exposure

Information was obtained from two schools that had used the *Saxon Phonics and Spelling* program in prior years. In particular, schools B and C had used the program during the prior year with students currently participating in the study. This provided an opportunity to examine whether or not there was an impact associated with years of usage. For example, did students already exposed to the program show greater or weaker benefits during the second year as compared to students in their first year of exposure? In order to address this question, multilevel analysis were conducted that included prior exposure.

It is important to note that complete data was not available on all students attending these schools. That is, for some students, teachers did not know whether or not they had been exposed to *Saxon Phonics and Spelling* during the prior year. In addition, prior usage information obtained indicated that the program was not used consistently or with the intensity that it was used during the study year. Indeed, results showed no significant differences among the groups, see Table E17. That is, prior exposure to the program did not have a significant impact on performance.

Table E17. Differences by Prior Exposure

	ITBS Word Analysis	ITBS Spelling	ITBS Reading Words	Ganske Spelling	M&P Spelling	Dolch Word List
Coefficient (Std. Err.)	5.99 (5.30)	4.65 (2.93)	14.17 (9.27)	-.12 (3.25)	-7.43 (6.06)	-2.40 (4.83)
t-ratio	1.129	1.589	1.529	-.038	-1.226	-.497
Sig	.26	.11	.13	.97	.23	.62

Analysis of Program Effects

Prior to discussing the results found, it is important to reiterate that there were a number of similarities between control and treatment classrooms. Both types of classrooms taught similar content such as phonemic awareness, fluency, spelling, vocabulary, writing, and so forth throughout the school year. There were also no major differences between treatment and control teachers in terms of how the lessons were structured. The only notable differences between these classrooms over the course of the school year was the 50-60 minute explicit phonics instruction treatment group students received via *Saxon Phonics and Spelling*. Given this information, prior research, and the fact that the duration of the study and exposure to the program occurred during *one* school year, small (.20) to moderate (.50) effect sizes were expected. It should also be noted that according to Slavin (1986), a leader in educational research, an effect size of .25 is considered educationally significant.

Independent Sample t-tests

Table E18 describes the means for the treatment and control groups for the outcome measures at post-testing. Independent samples t-tests were conducted for each of the key outcomes. Although Dolch Word scores were higher for treatment students than control students, it was not significant at the .03 level, a significance level that has been adjusted for multiple

comparisons. The multilevel models described in the next section take into account pretest scores and also incorporate dependency issues described above as a result of the hierarchical nature of the data.

Table E18. Post Test Scores for Students (Independent Samples t-test Results)

	N Control	N Treatment	Control Mean (SD)	Treatment Mean (SD)	t	Sig. value
ITBS – Word Analysis SS (1st and 2nd graders)	308	360	168.00 (23.82)	168.96 (25.85)	0.497	.62
ITBS – Reading Words SS (1st grade only)	163	161	151.47 (12.90)	153.17 (11.62)	1.243	.22
ITBS - Spelling SS (2nd grade only)	147	200	173.77 (15.22)	174.26 (13.97)	0.308	.76
Morris & Perney Spelling Test – PCT (1st grade only)	158	165	75.61 (23.62)	78.21 (18.68)	1.094	.28
Ganske Spelling Test – PCT (1st and 2nd graders)	297	361	63.86 (28.01)	65.30 (23.92)	0.705	.48
Dolch Words – PCT (1st and 2nd graders)	286	364	88.55 (20.73)	91.61 (16.20)	2.049	.04

Multilevel Models

The multilevel models were run for each of the key outcomes using a three-level formulation. In this formulation, the first level examines changes over time for each individual. The second level includes student level covariates. The third level incorporates teacher/school level information. The first set of initial models examines only the main effects of the program (see Appendix F for mathematical description of the model).

Separate multilevel models were run for each of the following outcomes:

- ITBS – Word Analysis Scale Score (1st and 2nd graders)
- ITBS – Reading Words Scale Score (1st grade only)
- ITBS - Spelling Scale Score (2nd grade only)
- Morris & Perney Spelling Test – Percent of Total (1st grade only)
- Ganske Spelling Test – Percent of Total (1st and 2nd graders)
- Dolch Words – Percent of Total (1st and 2nd graders)

Student level covariates in the model include:

- *Group* (Treatment=1; Control=0)
- Ethnicity (four category measure; categories modeled include *Hispanic, White*, *African American* and *Other* is the reference category)
- Gender (*Female*=1)
- *Grade* level

Other individual level covariates including special education status, Limited English Proficiency status, and free/reduced lunch were also available. However, due to small sample sizes (less than 61) and/or extensive missing data for these variables, these covariates were excluded from the multilevel analysis as this would significantly reduce the analytical sample. Teacher/school level covariates included in the model included school, teacher knowledge and preparation to teach the five elements of reading, spelling and writing, and teacher use of effective literacy practices.

The direct effects multilevel model was run on each of the measures noted above. Table E19 summarizes the results of the main program effects. Note that each measure in Table E19 corresponds to the program effect coefficients estimated for that dependent variable from a separate multilevel model. Significant differences (at the .03 level) were observed between the treatment and control groups at posttesting for the following measures: ITBS Word Analysis and ITBS Spelling. A marginally significant difference was observed for ITBS Reading Words (at the .07 level). Specifically, growth was greater in the treatment group as compared to the control group. It should also be noted that the coefficients corresponding to the pretest is not significant (indicating no significant difference between control and treatment) but the differences at posttesting is significant and positive (indicating that the treatment group is doing better).

Note that unlike the results presented in Table E18, these analyses incorporate important pretest data, student level demographic variables, and teacher/school information. When this is done via multilevel modeling, significant differences are obtained as described above. The effect sizes are also calculated; the effect sizes for the effect of the program on the ITBS subtests of reading words and spelling can be described as small, though educationally significant (d>.25).

Table E19. Main Program Effects from Three-Level Models

Outcome Measures	Coefficient	Std. Error	t-ratio	Sig. Level	Effect Size[49]
ITBS Word Analysis-Pretest	-0.759	2.440	-0.311	.76	---
ITBS Word Analysis-Posttest	3.967	1.831	2.166	.03*	.16
ITBS Reading Words-Pretest	-0.091	2.062	-0.044	.97	---
ITBS Reading Words-Posttest	3.417	1.910	1.789	.07**	.28
ITBS Spelling-Pretest	0.117	1.687	0.069	.95	---
ITBS Spelling-Posttest	4.178	1.835	2.277	.02*	.30
Morris & Perney Spelling-Pretest	2.886	4.971	0.581	.56	---
Morris & Perney Spelling-Posttest	-0.820	2.849	-0.288	.77	-.04
Ganske Spelling-Pretest	2.504	2.685	0.933	.35	---
Ganske Spelling-Posttest	-0.540	1.986	-0.277	.78	-.02
Dolch Word List-Pretest	6.383	3.231	1.975	.05**	---
Dolch Word List-Posttest	-2.081	2.190	-0.950	.34	-.11

*Significantly different after applying Sidak correction.**Marginally significant at the p<.07 level.

[49] Formula for calculating the effect size is in Appendix F.

Multilevel Models with Subgroup Effect[50]

Multilevel subgroup effects were analyzed for variables that had a sample size of 61 or greater[51]. The main effects three-level multilevel models were re-specified to re-estimate program effects for the following subgroups: gender (*female)*, grade (2nd), ethnicity (*Hispanic*, *African American, White* with *Other* as the reference category), free/reduced lunch status, and school (schools B, C, D, E, and F with school A as the reference category). First, a single model was implemented incorporating all of the interactions for the various subgroup effects. However, given strong correlations between the various interaction terms and multicollinearities in the model, a single model (with all of the interactions) was resulting in unstable estimates for a few of the interaction terms. Consequently, the subgroup effects were obtained by adding the interaction term(s) corresponding to each subgroup separately to the main effects model. Thus, separate models were run to obtain subgroup effects.

It is important to view this analysis as exploratory for a number of reasons: (i) the treatment and control groups were not randomized by subgroups; (ii) the sample sizes for a number of the subgroups are quite small; and (iii) differences were obtained between the treatment and control groups at baseline for some of the subgroups.

Tables E20-E24 summarize the results of the subgroup analyses for the following measures: ITBS Word Analysis, ITBS Reading Words (1st grade only), ITBS Spelling (2nd grade only), Ganske Spelling, Morris and Perney Spelling (1st grade only), and Dolch Word tests. Only statistical significant or marginally significant results are presented.

Significant effects were obtained for grade, ethnicity, school and free/reduced lunch status interaction terms for the ITBS Word Analysis test. Specifically, 2nd grade students, African Americans, students receiving free/reduced lunch status, and students attending schools D and F showed positive program effects. Statistically significant effects were also obtained for the gender, ethnicity, free/reduced lunch and school interaction terms for the ITBS Spelling test. Specifically, females, whites, students receiving free/reduced lunch, and students attending school D showed favorable program effects on this measure. Significant interactions were also observed between school and the ITBS Reading Words, Ganske Spelling, and Dolch Word tests. Note that the consistency in the observed effects for school D and F, and free/reduced lunch status across multiple measures means that we can be more confident in these results and suggests that the program produced positive effects among these groups of students.

[50] The interaction of group assignment and other subgroup classifications was also examined. However, due to the small sample available, these analyses were conduced via ANCOVA and is presented in the next section.

[51] This is in accordance with the procedures employed by NAEP researchers (Swinton et al., 2001). This number was obtained by determining the sample size necessary to detect an effect size of .5 and have a power of .8.

Table E20. Subgroup Effects from Three-Level Models-ITBS Word Analysis

	Grade	African Americans	School D	School F	Free/Reduced Lunch
Coefficient (Std. Err.)	4.61 (2.23)	5.74 (3.02)	14.97 (5.08)	9.50 (3.32)	8.71 (2.99)
t-ratio	2.065	1.897	2.949	2.859	2.910
Sig	.04	.06	.004	.005	.004

Table E21. Subgroup Effects from Three-Level Models–ITBS Reading Words

	School F
Coefficient (Std. Err.)	6.35 (2.90)
t-ratio	2.188
Sig	.03

Table E22. Subgroup Effects from Three-Level Models–ITBS Spelling

	Females	Whites	School D	Free/Reduced Lunch
Coefficient (Std. Err.)	4.64 (2.20)	5.38 (2.20)	15.41 (5.28)	7.31 (3.23)
t-ratio	2.105	2.444	2.917	2.263
Sig	.04	.02	.004	.02

Table E23. Subgroup Effects from Three-Level Models–Ganske Spelling

	School B	School D	School E	School F
Coefficient (Std. Err.)	9.82 (5.11)	25.43 (5.42)	9.23 (4.18)	9.88 (3.50)
t-ratio	1.924	4.693	2.206	2.821
Sig	.05	<.001	.03	.005

Table E24. Subgroup Effects from Three-Level Models–Dolch Word List

	School D
Coefficient (Std. Err.)	13.36 (6.64)
t-ratio	2.013
Sig	.04

ANCOVA with Subgroups

For subgroups with a sample size of 61 or smaller, (i.e., English language status, special education, and literacy level), exploratory ANCOVA analyses were performed. It is important to reiterate that due to the small sample sizes and likely dependency issues, no causal, conclusive statements should be made. These results are presented for preliminary, exploratory purposes. ANCOVA results showed no significant differences between treatment and control students who spoke a language other than English at home, were in special education, and who were at high or average literacy levels *after* controlling for pretest differences, as measured by all assessment measures. This means that there was no difference between treatment and control students in phonics and spelling skill levels at posttesting among these subgroups of students. However, among low literacy level students, there was a significant difference as measured by the Dolch Word test. Specifically, low performing treatment students had higher scores than low performing control students after equating the students on pretest scores, see Table E25.

Table E25. ANCOVA Results for Students of Low Literacy Ability

Measure		N	Mean*	Std. Error	F	df	Sig. value**
Dolch Word (1st and 2nd graders)	Control	46	69.90	3.88	4.836	1, 114	.03
	Treatment	68	80.97	3.19			

*Adjusted for pretest.

Appendix F

Mathematical Details of Multilevel Models and Statistics

The Structure of the Two-level Multilevel Model for Growth in Treatment Students

The model can be described with the following equations:

Level-1 Model

$$Y = P0 + P1*(Time) + E$$

Level-2 Model

$$P0 = B00 + R0$$
$$P1 = B10$$

Note that *Time* is the variable that differentiates between pre (0) and posttest (1).

The Structure of the Three-level Multilevel Model for Main Program Effects

Level 1 Model

The basic logic is to study if at pretest (time=1) there are pre-existing differences after controlling for a variety of measures. At level one, we focus on the trajectory of outcomes at the pretest and the posttest. The level one had the following function form:

$$Y = P0 + P1*(Time) + E$$

In the model above *P0* is a measure of the pretest, *P1* is a measure of change in outcome from pretest to posttest. *P0* and *P1* are calculated for each individual.

Level-2 Model

At level 2, we examine if *P0* and *P1* vary between individuals.

The level 2 model is:

$$P0 = B00 + B01*(GROUP) + B02*(FEMALE) + B03*(GRADE) + B04*(WHITE) + B05*(HISPANIC) + B06*(BLACK) + R0$$
$$P1 = B10 + B11*(GROUP) + B12*(FEMALE) + B13*(GRADE) + B14*(WHITE) + B15*(HISPANIC) + B16*(BLACK)$$

The key coefficients of interest include:

B01 is a measure of differences in outcomes between the treatment and control groups at pretest.

B11 is a measure of differences in outcomes between treatment and control groups at posttest.

Level-3 Model

The level 3 model is:

B00 = G000 + G001(SCHOOLA) + G002(SCHOOLB) + G003(SCHOOLC) + G004(SCHOOLD) + G005(SCHOOLE) + G006(TKNOWLDG) + G007(INSTRUCT) + U00
B01 = G010
B02 = G020
B03 = G030
B04 = G040
B05 = G050
B06 = G060
B10 = G100 + G101(SCHOOLA) + G102(SCHOOLB) + G103(SCHOOLC) + G104(SCHOOLD) + G105(SCHOOLE) + G106(TKNOWLDG) + G107(INSTRUCT)
B11 = G110
B12 = G120
B13 = G130
B14 = G140
B15 = G150
B16 = G160

Effect Size Calculations

The formula for calculating effect size from multilevel analysis is from Raudenbush et al. (2005):

$$\delta = \frac{\gamma_1}{\sqrt{\tau^2 + \sigma^2}}$$

where:

γ_1 is an estimate of program effect
τ^2 is the variation between clusters
σ^2 is the variation within clusters
$\tau^2 + \sigma^2$ is total variation.

The formula for calculating effect size from ANCOVA is obtained from Lipsey (1990). Eta^2 [i.e., proportion of variance accounted for (PV)] obtained from SPSS 14.0 was converted to Cohen's *d* to ease interpretation.

$$d = \sqrt{\frac{4\,(PV)}{1\text{-}(PV)}}$$

Adjustments Made for Multiple Comparisons

Sidak's correction for multiple comparisons was applied using the following rationale:

> If you test for the significance of a hypothesis using variables that are mutually correlated the Bonferroni correction is too conservative. For example, in an RCT a number of outcome variables are fully correlated. In that case knowledge of the outcome of a single test of a difference between the control and experimental group on a single variable, would be sufficient to know the outcome of the other tests on the other outcome variables. The usual Bonferroni correction would be way too conservative. In the case of correlated outcome variables a corrected alpha is required which is in between no correction at all and full, Bonferroni, correction. SISA [Simple Interactive Statistical Analysis software] allows you to add the mean correlation between variables as a parameter. For this you need the usual triangular matrix (without the diagonal) of the correlations between the outcome variables, sum the correlations and divide the result by the number of correlations used. A mean correlation of zero ('0') gives you full Bonferroni adjustment, a mean correlation of one no adjustment at all, for other values of the correlation you will get a corrected alpha which is in between the two extremes. (Rationale from Sankoh, A. J, Huque, M. F., & Dubey, S.D.,1997, para. 9).

To implement this adjustment, the Simple Interactive Statistical Analysis (Sankoh, Huque, & Dubey, 1997) was used. The results are as follows:

** Considering a correlation of 0.75 **
** To get an alpha level overall of 0.05 **

Sidak's adjustment
Lower the alpha for each test to 0.03
z-value for single sided testing: >=1.85
z-value for double sided testing: >=2.14

** Considering a correlation of 0.75 **
** To get an alpha level overall of 0.10 **

Sidak's adjustment
Lower the alpha for each test to 0.07
z-value for single sided testing: >=1.51
z-value for double sided testing: >=1.84