Assessing the Effects of the Reading Success Level A Program with Fourth-grade Students at a Title I Elementary School

Jaclyn M. Reed, Nancy E. Marchand-Martella,

Ronald C. Martella, and Russell L. Kolts

Eastern Washington University

Abstract

The purpose of this investigation was to assess the effects of the *Reading Success Level A* program on the comprehension skills of 93 fourth graders across four general education classrooms. Two general education teachers participated in this study over a 6-month period. Pre- and posttest data were collected on individual student performance using the *Scholastic Reading Inventory (SRI)*. In addition, within-program assessments including mastery quizzes and tests were administered as part of the program. Results showed that students who participated in *Reading Success Level A* demonstrated statistically significant gains in reading comprehension performance. In addition, at-risk readers made similar gains to those readers who were not at-risk indicating that *Reading Success Level A* was effective across students.

Without a doubt reading is the most important skill that students can acquire in school (Meese, 2001). It is closely tied to writing, spelling, mathematics, and content area activities. Reading at high levels is associated with continued academic success, significantly reduced risk for school dropout, and higher rates of entering college and finding successful employment (Lyon, 1999; Snow, Burns, & Griffin, 1998). Despite the importance of reading in our society, statistics continue to show that high percentages of students struggle with reading. For example, the National Center for Education Statistics (NCES) found that 40% of fourth graders and 32% of eighth graders did not meet the basic requirements set forth by the National Assessment of Educational Progress in reading (NCES, 2003). These basic requirements include skills in examining literature, reading for information, and deciphering specific information to perform tasks. Furthermore, reading is the area where most students qualify for special education services (Meese, 2001).

Correspondence to Nancy Marchand-Martella, Ph.D., Department of CEDP, Martin Hall 135, Eastern Washington University, Cheney, WA 99004; e-mail: marchandmartella@sisna.com.

Given the high percentage of students who struggle in reading and the debate over how best to teach reading, Congress mandated the formation of the National Reading Panel (NRP) in 1997. The NRP's task was to review and evaluate supporting literature on effective reading instruction and to summarize the results; over 100,000 studies published in reading since 1966 and 15,000 before that time were examined. The NRP Report (National Institute for Child Health and Human Development [NICHD], 2000) found five areas of instruction critical to the success of beginning readers and those in need of remediation. These areas include: phonemic awareness, phonics, fluency, vocabulary, and text comprehension. Each of these five areas of effective reading instruction is best acquired through systematic and explicit instruction. Systematic instruction is defined as "the plan of instruction that includes a carefully selected set of letter-sound relationships that are organized into a logical sequence" (Armbruster, Lehr, & Osborn, 2003, p. 19). Explicit instruction pertains to "a systematic method of teaching with emphasis on proceeding in small steps, checking for student understanding, and achieving active and successful participation by all students" (Rosenshine, 1987, p. 34). When reading instruction is systematic and explicit, reading skills are acquired at higher levels (Armbruster et al., 2003).

The first three areas of effective reading instruction (phonemic awareness, phonics, and fluency) are typically referred to as "learning to read" or decoding skills (Schieffer, Marchand-Martella, Martella, & Simonsen, 2002). Emphasis for beginning readers is on sound-symbol correspondence, blending and reading words the "fast way," and fluent text reading. When "learning to read" skills are well developed, comprehension is enhanced (Osborn, Lehr, & Hiebert, 2003). According to Armbruster et al. (2003), increased reading fluency "frees students to understand (comprehend) what they read" (p. 31).

In addition to "learning to read" skills, it is important that systematic and explicit instruction be provided when acquiring "reading to learn" skills. "Reading to learn" focuses on vocabulary building activities and text comprehension strategies to help students gather meaning from text. Most vocabulary is learned indirectly through conversations with others, listening to others read, and extensive independent reading (Armbruster et al., 2003). In fact, Cunningham and Stanovich (1998) found large differences in student vocabulary based solely on how much independent reading they participated in outside of school. During this study, the authors analyzed how many rare words per 1,000 appeared in selections of major sources of spoken and written language such as newspapers, textbooks, television, and adult speech. They then compared students' exposure to rare words to

their vocabulary repertoire. Exposure to rare words was found to be an indicator for vocabulary growth and development. Further, Cunningham and Stanovich (1998) found that reading volume accounted for substantial differences in vocabulary development (i.e., the more children read, the more vocabulary they are exposed to). Because many vocabulary words are not learned through indirect means such as exposure to spoken and written language and reading books, direct vocabulary instruction, with a focus on specific words and word learning strategies, has been targeted as important for instruction (NICHD, 2000).

In addition to vocabulary building, text comprehension strategies are needed to enhance "reading to learn" skills. Text comprehension strategies are "conscious plans--sets of steps that good readers use to make sense of text" (Armbruster et al., 2003, p. 49). Taken together, instruction that focuses on vocabulary building and text comprehension is critical for student success and can collectively be referred to as "comprehension skills." Comprehension skills are viewed as the very essence of reading (Durkin, 1993). These skills allow students to read for information and for entertainment, translating reading into a purposeful, life-long activity (NICHD, 2000).

Some educators believe that readers who read fluently will automatically comprehend what they read. Unfortunately, this statement is not necessarily true—there is a higher likelihood that fluent students will understand what they read; however, research shows that even those students who have mastered fluent decoding skills early can struggle to comprehend in later grades (Snow, 2002). In fact, comprehension of text involves a complex interaction between the reader's individual skills, the features of the text being read, and the activity of processing the text through decoding, semantic processing, and self-monitoring (Snow, 2002). Therefore, students must be taught "to use specific cognitive strategies or to reason strategically when they encounter(ed) barriers to comprehension in reading" (NICHD, 2000, p. 5). Interestingly, more than 20 years ago, Chall (1983) found that around the fourth grade, one would often see "word callers," students who were able to sound out and read words accurately and fluently but with little comprehension when words appeared in connected text. Chall called this phenomenon the "fourth grade slump." Text at this grade level becomes less narrative and familiar and more expository in nature; thus, the complexity of text increases, requiring students to think about and understand what they read (Chall, 1983; Dymock, 1993; Fournier & Graves, 2002; Snow et al., 1998). Given this phenomenon, it seems critical that systematic and explicit instruction, particularly before or at least during the fourth grade, is needed-

(Cunningham & Stanovich, 1997; Lyon & Moats, 1997).

When systematic and explicit instruction in text comprehension is provided, it should include comprehension monitoring (using comprehension strategies such as re-reading difficult text while reading), graphic and semantic organizers, instruction on story structure to understand how and why a passage is constructed in a particular way, question answering and question generation to aid in drawing inferences, summarization of text, and the use of multiple-strategy teaching techniques (e.g., reciprocal teaching) that emphasize continued teacher-student interaction during instruction (Armbruster et al., 2003; NICHD, 2000). These comprehension strategies help students gather meaning from text. Similarly, Lyon (1998) identified skills in summarizing, predicting, clarifying, and generating questions as primary indicators of good comprehenders. Additional research substantiates the importance of these strategies in students' acquisition of meaning from text, awareness while reading, and prediction of future reading success (Armbruster et al., 2003; Cross & Paris, 1988; Dymock, 1993; Lyon & Moats, 1997).

One program that teaches text comprehension skills, with additional practice in vocabulary building, is *Reading Success Level A* (Dixon, Klau, Rosoff, & Conrad, 2002). *Reading Success Level A* is a systematic and explicit supplemental program specifically designed to teach students skills such as identifying main idea, making inferences, paraphrasing, and acquiring meaning from figurative language, including the explicit teaching of key vocabulary words such as *anaphora*, *expository*, and *paraphrase* to help achieve reading success. The program involves track sequencing rather than unit sequencing. Thus, concepts learned previously are incorporated with new information throughout the program. Scaffolding and review are central tenants to *Reading Success Level A* (fourth-grade), *B* (fifth grade), and *C* (sixth grade). No published investigations were found on the effects of any level of *Reading Success.*

Research in the area of reading demonstrates the effectiveness of explicit and systematic instruction in vocabulary and text comprehension strategies. In fact, "good instruction is the most powerful means of promoting the development of proficient comprehenders and preventing reading comprehension problems" (Snow, 2002, p. xvii). The purpose of this investigation was to assess the effects of the *Reading Success Level A* program on the comprehension skills of fourth graders across four general education classrooms.

Method

Participants

This investigation began with all fourth-grade students across four general education classrooms at one elementary school. Two fourth-grade general education teachers provided instruction using the *Reading Success Level A* program.

Students. Out of the 95 total fourth-grade students, 14 students did not begin the program with a fourth-grade reading level as determined by the *Scholastic Reading Inventory (SRI)* (Scholastic, 1999). Two of these 14 students had an Individualized Education Program (IEP) and were diagnosed with a Conduct Disorder. In addition to *Reading Success Level A,* the 12 students who did not read at grade-level (excluding the two students with IEPs) participated in *Corrective Reading Decoding Level B2* (Engelmann et al., 1998) or *Read Naturally* (Ihnot, 1991) as part of Learning Assistance Program (LAP) tutoring. Additionally, one of the 14 students received English Language Learner (ELL) special services, in addition to LAP tutoring, during the study. One of the students with an IEP was dropped from the program due to a high rate of absenteeism and another student moved during the course of the investigation (received LAP tutoring). Both were excluded from the study, leaving a total of 93 students (1 with an IEP, 10 receiving LAP tutoring only, and 1 receiving LAP tutoring and ELL special services).

The 93 students who participated in the program ranged from 9 to 10 years of age. Of these students, 47 were females and 46 were males; 93.5% were Caucasian (n = 87), 2.2% were Hispanic (n = 2), 1.1 were Asian (n = 1), 1.1% were African American (n = 1), 1.1% were Russian (n = 1), and 1.1% were American Indian or Alaskan Native (n = 1). Approximately 13% (n = 12) of the students required some form of special services for part of the school day. Students were not skill grouped for the reading comprehension program; they received whole class instruction as they did with all other academic subjects.

Teachers and research assistants. Two general education teachers provided instruction. Teacher A had 26 years of teaching experience and had been involved in creating district and school-wide benchmarks in reading for many years. She served on the District Learning Team and was the district's Language Arts Representative. Teacher B had 20 years of teaching experience, 14 in special education and 6 in general education (grades preschool through sixth). He also had advanced training and expertise in Direct Instruction, having taught various programs including *Reading Mastery* (Engelmann & Bruner, 2003) and *Corrective Reading* for approximately 14 years, studying

at the Slingerland Institute and the Eugene Summer Institute. Both teachers had a Master's degree in curriculum and administration. Each teacher co-taught with another fourth-grade teacher. Teachers A and B taught reading and language arts where their co-teaching counterparts (Teachers C and D) taught science and mathematics. Teacher A instructed students from Classrooms A and C. Teacher B instructed students from Classrooms B and D.

In addition, a graduate student in school psychology (first author) and a post-baccalaureate student in special education assisted in re-teaching lessons, administering mastery quizzes and test re-takes, and scoring mastery quizzes and tests. The graduate student had a Bachelor's degree in applied psychology with an emphasis in child and adolescent development. She had finished her first quarter of graduate studies in school psychology when this investigation began and participated in the study to fulfill her academic research project requirement. The post-baccalaureate student was selected to assist the graduate student in scoring and reteaching. This student had a Bachelor's degree in English and was working toward a primary endorsement in special education. He was also working as a substitute teacher in a local school district.

Setting

This study took place in a Title I urban elementary school located in the Pacific Northwest. Out of the 601 students enrolled in the school, approximately 36.4% qualified for free- and reduced-price meals. Fourth-grade scores in reading and writing on the *Washington Assessment of Student Learning* (*WASL*), during the 2002-2003 school year, were 89.7% and 80.4%, respectively. In 2002-2003, district averages were 79.1% and 69% in reading and writing, respectively; state averages were 66.7% and 53.6% for these same areas.

Targeted Curriculum

Reading Success Level A (Dixon et al., 2002) was implemented in this investigation. This was the first year of program implementation; in previous years, the teachers taught comprehension strategies informally without the use of a published program.

Reading Success Level A introduces comprehension concepts and continually reviews subsequent concepts learned in the program. The teacher's guide includes 80 lessons with a 2-in (5.08 cm) left-hand column including a general script, hints on how and when to introduce additional examples, and additional word meanings. It also includes the answers to all of the quizzes and tests. Student workbooks are designed so students can follow along as the teacher instructs the lesson

and do not include quizzes or tests. A set of blackline masters of all quizzes and tests is provided. Each quiz and test comes with a blackline score sheet.

The program teaches comprehension strategies using explicit and systematic instruction (Dixon et al., 2002). Teachers demonstrate these strategies (modeling), conduct scaffolded practice so students can apply these strategies (guided practice), and include application of these strategies over time with a variety of examples (independent practice and review). The concepts taught in this program include anaphora and classification, main idea, inference and literal, fact and opinion, author's purpose, paraphrase, word meanings, figurative language, and bonus terminology.

Anaphora and classification. Students learn to use anaphora, a "pronoun or other words used to refer to some other word or name" (Dixon et al., 2002, pp. 2-3) and classification of details to choose between good and poor main idea statements.

Main idea. Students are initially given short passages and are asked to determine who or what is talked about the most, classify selected details from the passage, and fill in a main idea box consisting of the two concepts to derive their main idea statement. As the lessons progress, the main idea box is faded and students are asked to write their own main idea statement.

Inference and literal. Students are exposed to many different kinds of inferred and literal questions including inferences about characters, setting, and prediction and are then explicitly taught how to recognize the difference between a literal or inferred question.

Fact and opinion. Students learn the difference between fact and opinion by reading short passages and determining if the details they are reading can be proven with dictionaries or encyclopedias. This skill is continually practiced throughout subsequent lessons.

Author's purpose. By using a simple checklist, students are taught to identify whether an author is writing to entertain, inform, or persuade. Eventually the checklist is faded and the students are asked to determine author's purpose by reading the passage.

Paraphrase. Students are asked to paraphrase sentences by either changing the word order or replacing certain words that mean the same thing. Eventually students paraphrase whole passages and rewrite passages in their own words.

Word meanings. Students are also asked to determine the use of a word in different contexts. For example, a word like "change" has several meanings. The particular meaning of "change" must then be determined by the context of the passage.

Figurative language. Students are introduced to poetry and the use of literal and figurative language. Gradually the students are

asked to not only determine if the language is figurative but what the figurative words used in a given poem or passage mean.

Bonus terminology. Every five lessons, during a quiz or test, a bonus term is introduced and students are expected to know the spelling of the word, the meaning of the word, the derivation of the word, and the current usage of the word.

Quizzes/tests. After every five lessons a built-in mastery quiz is administered; every 20 lessons a comprehensive mastery test is conducted. *Reading Success Level A* does not provide a placement test.

Core/Additional Curricula

In addition, all fourth graders participated in the reading motivational program, *Scholastic Reading Counts!* (Scholastic, 2002). *Scholastic Reading Counts!* uses student Lexile scores (Stenner, Smith, Horabin, & Smith, 1988) to determine grade level appropriate books. Students read books and take computerized comprehension quizzes. They must pass these quizzes with 80% accuracy to move on and have the number of words for the "passed" book added to their total number of words read for the year. Alternate forms are provided for each quiz trial to prevent practice effects.

The 12 students who were involved in LAP tutoring with or without ELL special services also participated in *Read Naturally* or *Corrective Reading Decoding Level B2* before school. *Read Naturally* combines teacher modeling, repeated reading, and progress monitoring to improve reading fluency. *Corrective Reading Decoding B2* is an explicit and systematic remedial reading program.

Dependent Variables and Measures

All students were assessed before and after program implementation using the *SRI.* The number of words correct per minute (WCPM) was obtained using the *Read Naturally* program. Student progress on within-program assessments was also recorded. The total number of words read by the end of the year was tracked via the *Scholastic Reading Counts!* program. Additionally, scores on the *WASL* were gathered. Finally, a social validation survey was provided.

SRI. Students were assessed before and after the *Reading Success Level A* program using the *SRI*. The *SRI* is a computer-adaptive assessment used to determine how well students read and comprehend literature and expository text at varying difficulties. The *SRI* focuses on comprehension skills including identifying details in a passage, identifying cause-and-effect relationships and sequencing of events, drawing conclusions, and making comparisons and generalizations. The *SRI* includes authentic text taken from magazines, newspapers,

textbooks, and fiction. Based upon the students' answers as they are taking the test, the computer moves to either easier or more difficult questions. The *SRI* provides a Lexile for each student. The Lexile score measures student performance within a range of Beginning Reader (BR) to 1700+. Readers are given a score and their reading level range is determined by adding 50 and subtracting 100 (e.g., *SRI* Lexile = 1200; reading level range = 1100 – 1250). During this study, the Lexile, normal curve equivalents (NCE), percentile ranks, and stanines were calculated before and after instruction was provided. NCE's have a mean of 50 and standard deviation of 21.06**.**

WCPM. WCPM data were gathered during the spring of the fourth-grade year. These data were retrieved from the school's *Read Naturally* program. Throughout the year each fourth-grade student was required to take *Read Naturally* quizzes periodically that measured reading fluency by documenting speed and accuracy of words read within a 1-min interval. The last WCPM score of the school year was correlated with beginning and ending *SRI* stanine, percentile rank, and *SRI* Lexiles to determine if WCPM (fluency) affected reading comprehension.

Within-program assessments. Mastery quizzes and tests were included with the *Reading Success Level A* program. Quizzes consisted of approximately four to five multiple choice questions plus two bonus questions and appeared every five lessons. The graduate student scored each quiz and test and recorded each student's score on the score sheets included with the program.

Tests consisted of approximately 12-20 multiple choice and short answer questions plus two bonus questions. Tests appeared every 20 lessons. These tests were used to evaluate student performance in the same manner as the quizzes. Tests were more comprehensive than the quizzes but the same re-testing procedures applied.

Each quiz/test was scored as number correct out of 100, converted to a percentage, and reported as percentage correct. Bonus questions were not reflected in each student's score, however, these questions were corrected and re-taught, if necessary.

Standards for re-taking quizzes and tests were established in this study. Each student was expected to reach 80% mastery on each quiz and test. Individual students who did not reach 80% mastery were re-taught the concepts of the quiz or test explicitly by the research assistants and were given two additional attempts at reaching mastery. If students did not reach mastery after all attempts were made, they moved on in the program with the rest of their classmates. Re-takes generally occurred one to two times per week as needed. The total number of retakes per student was recorded and analyzed.

WASL. WASL scores in reading and writing from the previous fourth-grade year (2002-2003) were compared to *WASL* scores in reading and writing for fourth graders in this study. Comparisons were made between scores across these 2 years.

Social validation survey. The social validation survey consisted of 10 questions and was administered to teachers at the end of the program. This survey was developed by the first author, along with two faculty involved in the research project, to address issues about the program and its implementation. Questions 1, 3, and 8 asked teachers to rate the program on a Likert Scale (1 = very poor to 5 = excellent). Question 1 asked about the adequacy of the training for the program. Question 3 asked if the teachers received adequate support while implementing the program. Question 8 asked the teachers to rate how easy it was to implement the program. Questions 2, 4, 5, 6, 7, 9, and 10 required short answers and comments on how the program could be improved, if differences were seen in students' comprehension, and whether or not the program should be implemented in the future.

Design and Procedures

A pre-experimental research design was used during this study (Martella, Nelson, & Marchand-Martella, 1999) for program evaluation purposes. All teachers and the graduate student in school psychology received training before program implementation.

Teacher and research assistant training. A representative from Classical Learning Universe conducted training at the school in one 4-hr session. He had a masters degree in special education, was a current special education teacher, and served as a national reading consultant. Training consisted of an explanation of the different skills learned in the program, the reasons for explicit instruction, the materials needed for the program, and the testing structure. Examples of how to provide explicit instruction were given as well as instruction on how to use error correction procedures within the program. Four teachers attended the training (the two teachers who provided instruction in this study and their co-teaching counterparts) as well as the graduate student. In addition, half way through the program a post-baccalaureate student at the local university was trained by the graduate student to give re-take exams, re-teach concepts, and score quizzes and tests during the program.

Reading Success Level A instruction. Instructional sessions were conducted approximately 4 days per week (Monday-Thursday) unless holidays or other school-related activities occurred. Students received instruction as a whole class. All 80 lessons were completed with an average duration of 20 min per lesson. The format of instruction

followed the outlined format of the program. Students completed items in the student workbook according to program guidelines. Instruction followed the same guidelines. The teachers monitored answers to questions in the student workbook while they taught the lesson. Teachers used choral and individual responding throughout each lesson. A considerable amount of guided practice was used in each lesson in accordance with the teaching guide script and lesson design. Both teachers provided additional examples and explanations for concepts learned and explained new vocabulary when needed. The *Reading Success Level A* program also emphasizes the importance of proper error correction procedures. It is expected that teachers would use a model, lead, test, re-test approach for error correcting. This approach includes the teacher first modeling a concept or use of a strategy explicitly. The teacher would then provide guided practice while the students attempted to use the strategy. The teacher would then move onto independent practice when a concept was mastered. Review and additional practice would be given as needed.

An example of teaching the concept for main idea from Lesson 9 follows. The teacher might begin by saying, "Class, we will be continuing to learn how to choose a good main idea statement today. Remember main idea has five steps. The five steps to main idea are. Get ready." (point or clap for choral responding). The class would then respond by repeating the five steps of main idea ([1] decide who or what is talked about most, [2] write a list of details, [3] classify the details, [4] fill in the main idea boxes, and [5] write a good main idea statement). Responses could be given through choral responding or individual responding. If the class or individual seemed unsure or unclear about one of the steps, the teacher would go over the steps again clearly. For instance, if the class missed the second step (i.e., write a list of details), the teacher might continue as follows, "Stop. My turn. Remember the five steps to finding main idea are decide who or what is talked about most, write a list of details (stated with emphasis), classify the details, fill in the main idea boxes, and then write a good main idea statement. When finding main idea first you decide who or what is talked about most and then you . . .everybody. Get ready." (point or clap for choral responding). The class would then respond, "Write a list of details." The teacher would continue by giving positive statements such as, "Good job identifying the second step to main idea, write a list of details," or by following the error correction procedure again until the class understands each step clearly. As students progressed through each lesson, the teacher would walk around and monitor students for understanding. If students continued to struggle with the concept, the teacher would stop the class and go over the

steps again. It is suggested within the *Reading Success Level A* program to post steps in the classroom so that students can remind themselves on a regular basis. During this study, both teachers posted steps in some manner within the classroom.

In compliance with the recommendations of the program, if 80% or more of the class missed a concept either on a quiz or test or during a lesson, the concept was re-taught by the classroom teachers to the whole class before subsequent lessons convened. The score sheets included within the program indicated the concepts on which the students struggled. Lessons were re-taught by the research assistants if 20% or less of all students failed to reach 80% mastery on a quiz/test. After re-teaching, the same quiz or test was re-administered. Re-teaching to the whole class occurred on four occasions during the 80-lesson program.

Program Fidelity

To ensure instructional fidelity across the 80 lessons of *Reading Success Level A*, observations of Teachers A and B were made. Each teacher was observed four times by the graduate student. An instructional fidelity checklist was developed by the graduate student and two faculty involved in the research project based on the elements emphasized in the program. The teachers involved in the program provided input and were given instruction on each of the points covered on the checklist. Teachers were rated from 0-5 (0= does not cover point at all during the lesson to 5= covers point well during the lesson) on five criteria. These criteria were: (a) Teacher follows format outlined by *Reading Success Level A* program; (b) Teacher often uses specific praise statements and provides immediate feedback; (c) Teacher monitors student responses frequently during the lesson; (d) Teacher re-teaches either part or all of a lesson(s) when needed and provides alternative or additional explanation(s) when needed; and (e) Teacher uses proper error correction procedures established by *Reading Success Level A* program.

In addition, under each of the major criteria specific sub-criteria were listed for clarification of each point. For example under criteria #5, "teacher models correct response and has students repeat task when error correcting" and "teacher uses proper amount of repetition and practice to establish mastery of a concept when errors have occurred" are included as sub-criteria. Based on these observations, teachers were given immediate feedback after the lesson during a brief meeting. They also received a copy of the completed observation checklist. Teacher A had a mean score of 4.0 while Teacher B had a mean score of 5.0 on each of the criteria. Scores on the fidelity checklist ranged from 3.5 – 5.0 across teachers and observations.

Interobserver Agreement

The graduate student set criteria for scoring subjective questions by giving examples and non-examples of acceptable responses. The teacher's guide was consulted for all other correct answers. The graduate student regularly double-checked scoring by the post-baccalaureate student to ensure agreement on scoring. Approximately every 20 student quizzes and tests were checked for scoring criteria. Scoring criteria consisted of the graduate student and the post-baccalaureate student both scoring a quiz or test and comparing each against the other. Agreement on scoring was calculated by adding the number of agreements and the number of disagreements and dividing by the number of agreements. The graduate student and the post-baccalaureate student had 100% agreement on scoring during this study.

Statistical Analyses and Comparisons

Comparisons were made on six variables. At-risk students were compared to non-risk students to determine whether the program was more beneficial to one group. Differences were also considered for gender. In addition, comparisons between performance of students receiving instruction from Teacher A versus Teacher B were made. Spring WCPM were also evaluated in comparison to national norms. Number of retakes per student was explored to evaluate the correlation between retakes and overall performance. Finally, *WASL* scores for 2002-03 were compared to *WASL* scores in 2003-04.

SRI. Non-risk students were compared to at-risk students on pre- and posttest Lexile scores, percentile ranks, and stanines. Students were considered at-risk if they scored 610L or below on the *SRI* pretest or if they participated in LAP tutoring or had an active IEP. This comparison was done using a Wilcoxon Signed-Rank Test. A Mann Whitney-U Test was also done to determine if the non-risk and at-risk groups differed in the extent to which they benefited from the program in a statistically significant way.

Comparisons were also made by gender. Males were compared to females on pre- and posttest Lexile scores, percentile ranks, and stanines. This comparison was done again using a Mann Whitney-U Test to determine change differences.

Finally, comparisons were made across teachers (i.e., Teacher A and Teacher B). Pre- and posttest Lexile scores, percentile ranks, and stanines for the two classrooms were compared to find if either group benefited more from the program. Specifically, differences in scores between Teacher A's classrooms (A and C) and Teacher B's classrooms (B and D) were tested for using a Mann Whitney-U Test.

WCPM. Students were compared to national norms on WCPM.

Comparisons were made based on WCPM data obtained from the *Read Naturally* program gathered in the spring of the 2003-04 school year. National norms were drawn from a sample of approximately 6,000 fourth-grade students in five mid-western and western states (Tindal, Hasbrouck, & Jones, 2005). The analysis used to compute this comparison was a z-test.

Within-program assessments. Total number of retakes was counted for each student. Total number of retakes was then examined in relation to risk, reading functioning, WCPM, and total words read for the year. A correlational analysis was done on this measure.

To explore relationships between retake frequency and outcome variables, correlations were calculated between number of total retakes per student and indices of reading function. These indices are defined as *Scholastic Reading Counts!* number of words read for the 2003-04 school year, spring 2004 WCPM, *SRI* Lexile scores (pre – and posttest), percentile ranks (beginning and ending), and stanines (beginning and ending).

WASL. The overall percentage of fourth-grade students meeting the *WASL* standard for 2002-03 was compared to percentages for 2003-04 in reading and writing. In addition, changes in percentage of fourth-grade students passing each *WASL* level were compared (Level 4 = above standard, Level 3 = met standard, Level 2 = below standard, and Level 1 = well below standard).

Results

SRI

Three analyses were conducted based on *SRI* scores—non-risk versus at-risk, gender, and Teacher A versus Teacher B.

Non-risk versus at-risk. A Wilcoxon signed-rank test was utilized to test for improvement between groups due to the various dependent variables. The results of the test revealed statistically significant differences between pre- and posttest scores on the SRI for both at-risk and non-risk groups (z_{NR} = -6.18, p < .001, n_{NR} = 63; z_{AR} = -4.54, p < .001, n_{AR} = 27) (See Table 1). In addition, the Wilcoxon signed-rank test revealed statistically significant changes on the measure of percentile rank and stanine from pre- to posttest for both groups (Percentile Rank: z_{NR} = -6.21, p < .001; z_{AR} = -4.71, p < .001; Stanine: z_{NR} = -5.56, p < .001; z_{AR} = -4.60, p < .001).

Additionally, results of a Mann-Whitney U test revealed statistically significant differences in the extent to which the groups benefited from the program with the at-risk group showing larger gains on all three measures (*SRI*, Percentile Rank, and Stanine), (*SRI*: U = 567.50, p < .013; Percentile Rank: U = 486.00, p < .001; Stanine: U = 605.00, p < .006).

Table 1
Summarization of Non-Risk vs. At-Risk SRI, Percentile Rank, and
Stannine Pre- and Posttest Scores

Measures	N	Pretest Score	Posttest Score	Change	z	p
Non-Risk						
SRI	63	839.38	985.78	+146.40	-6.18	<.001
Percentile Rank	64	69.27	84.38	+15.11	-6.21	<.001
Stanine	64	6.34	7.45	+1.11	-5.56	<.001
At-Risk						
SRI	27	509.96	750.41	+240.45	-4.54	<.001
Percentile Rank	29	25.48	57.34	+31.86	-4.71	<.001
Stanine	29	3.59	5.59	+2.00	-4.60	<.001

Gender. A Wilcoxon signed-rank test revealed no statistically significant differences in pre-post change on measures of *SRI*, percentile rank, and stanine between male and female students (*SRI*: U = 972.00, *p* < .75; Percentile Rank: U = 1015.00, *p* < .61; Stanine: U = 1045.50, *p* < .78). These results indicated that statistically significant changes on these three measures cannot be attributed to gender.

Teacher A versus teacher B. A Mann-Whitney U test was also performed to determine if statistically significant differences on *SRI* change, percentile rank change, and stanine change from pre- to post-test were seen between teachers implementing the program. Statistically significant differences on this measure would indicate that the program worked better for particular students based upon who was teaching the program, not the program itself. The Mann-Whitney U test revealed no statistically significant differences between teachers on any of the three measures. Mean ranks between classes for Teacher A and Teacher B on the measure of *SRI* change (M$_A$ = 46.59; M$_B$ = 44.36) were not significantly different, z = -4.04, *p* < .687. Mean ranks between Teacher A and Teacher B on the measure of percentile rank change (M$_A$ = 47.74; M$_B$ = 46.28) were not significantly different, z = -.261, *p* < .794. Finally, mean ranks between Teacher A and Teacher B on the measure of stanine change (M$_A$ = 46.91; M$_B$ = 47.09) were also not significantly different, z = -.032, *p* < .975.

WCPM

An independent samples z-test revealed that the fourth-grade students' WCPM at posttesting (m = 138.01, SD = 21.987) differed significantly from the national WCPM norms (m = 124.81, SD = 43.26), at the 50th percentile, for spring fourth-grade year on a significant level,

z = 2.94, p < .01. An independent sample t-test was also computed to compare the 50th percentile norms to the program sample score, t(.92) = 6.61, p < .001. The results of these tests revealed that the difference between the spring WCPM levels of this study and those established by national norms were statistically significant (see Table 2).

Table 2

Difference between *Reading Success Level A* **Mean Spring Fourth-Grade WCPM and Mean WCPM form 2005 National Normative Sample at the 50th Percentile**

Measure	m	SD	z	p	t	df	p
Reading Success Level A WCPM Spring (Fourth grade)	138.01	21.987	2.94	<.01	6.61	.92	<.001
National Standard WCPM Spring (Fourth grade)*	124.81**	43.26	—	—			

* Reference: Tindal, Hasbrouck, & Jones (2005).
** Mean from the normative sample at the 50th percentile

Within-Program Assessments

An independent samples *t* test was also used to test for differences between the at-risk group (m = 6.62) and no-risk (m = 3.56) group in regard to number of re-takes. As expected, this test revealed that those students in the at-risk group took more re-takes than those in the non-risk group, *t* (91) = -4.33, *p* < .001.

Correlations were calculated between number of total retakes per student and indices of reading function to explore relationships between retake frequency and outcome variables. Negative correlations were found on each of these indices of reading, indicating that as number of retakes increased reading level on each measure decreased. These correlations are to be expected since students with a lower reading level would be expected to need more practice time (re-takes) to understand a concept. It was found that number of re-takes was unrelated to change on indices of percentile rank and stanine with the exception of change in *SRI* score from pre- to posttest (r = -.229, *p* = .03). The negative correlation found between number of re-takes and *SRI* change indicates that the more re-takes students took, the less improvement they made on the *SRI* over the course of the study. Table 3 shows correlational data for all indices of reading functioning.

In addition, an independent samples *t* test was used to test for differences between males and females in regard to number of re-takes, revealing no statistically significant differences on measure of gender, $t(91) = 1.13$, $p = .261$.

WASL

The percentage of fourth-grade students who passed the WASL in reading and writing in the year 2002-03 was compared to fourth-grade students who passed the same standard in 2003-04. Goodness-of-fit (chi-square) analyses were performed on the data to determine if the difference on each WASL measure (reading and writing) between years was statistically significant. There were no statistically significant differences on rates of passing the WASL reading measure between years, $\chi2(1, N = 193) = 1.046$, $p = .307$, however, statistically significant differences were found on the WASL writing measure, $\chi2(1, N = 192) = 7.465$, $p < .006$. These results indicated that students exposed to the Reading Success Level A program passed the Washington State standard in writing at a significantly higher rate than those students taking it the previous year. A statistically significant difference was not found between years on the measure of WASL reading; however, the percentage of students passing the reading section for both years was above the state average. The percentage of students passing reading in 2002-03 and 2003-04 was 89% and 94%, respectively.

Social Validation Survey

On the social validation survey the teachers rated the program training average (3.0) to good (4.0) on a 5-point Likert scale, where 1 was very poor, 3 was average, and 5 was excellent (mean = 3.75). All of the teachers agreed that the support they received during the implementation of the program was excellent (mean = 5.0). The teachers rated the program as being fairly easy to implement on a daily basis (mean = 3.5, range = 3-4, where 1 was very difficult and 5 was very easy). Finally, the teachers were asked open-ended questions regarding the program. When asked what they liked about the program, all teachers agreed that the program was clear and easy to teach. All teachers agreed that this program transferred to other subject areas such as mathematics, science, and writing and that support of the program was "stellar." Some suggestions were made to increase massed practice activities within the program, to provide more tips for the teachers in the margins of the teacher's manual, and to provide some video training of the lessons being taught. All of the teachers noted significant improvement in the reading and writing skills of their fourth graders and felt strongly about using this program in their classroom again.

Table 3
Coorrelational Data on All Indices of Reading Functioning

	Reading Counts	WCPM	SRI (pretest)	SRI (change)	Percentile (begin)	Percentile (change)	Stanine (begin)	Stanine (change)
# of Total Retakes	-.559**	-.289**	-.422**	-.229*	-.490**	.018	-.497**	-.106
Reading Counts	—	.350**	.455**	.184	.473**	.082	.490**	-.052
WCPM	.350**	—	.446**	-.071	.506**	.179	.529**	.127
SRI (pretest)	.455**	.446**	—	-.382**	.965**	.602**	.948**	.358**
SRI (change)	.184	-.071	-.382**	—	-.325**	-.853**	-.290**	-.827**
Percentile (begin)	.473**	.506**	.965**	-.325**	—	.582**	.973**	.379**
Percentile (change)	.082	.179	.602**	-.853**	.582**	—	.539**	.856**
Stanine (begin)	.490**	.529**	.948**	-.290**	.973**	.539**	—	.406**

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Discussion

This study examined the effectiveness of *Reading Success Level A* (Dixon et al., 2002) on the reading performance of fourth-grade general education students. Results of the statistical analyses comparing non-risk and at-risk learners revealed statistically significant gains from pre- to posttesting on all indices of reading function. These significant findings indicate that at-risk learners benefited from the program at least as much, if not more, than non-risk learners. Students also performed significantly higher on posttest number of words correct per minute than the national normative sample at the 50th percentile, indicating students read as accurately or better than 50 out of 100 of their peers. The mean posttest WCPM (mean = 138) for students in the study was also higher than the minimum required by the Washington State Essential Academic Learning Requirements (EALRs) for spring of fourth-grade year (WCPM = 115-125).

In addition, statistically significant differences were found for the percentage of students passing in the area of writing on the Washington Assessment of Student Learners (WASL) from the previous year. Although differences on the percentage of students passing in the area of reading on the WASL were not found to be statistically significant, percentage of students passing in this area did increase (2002-03: 89%; 2003-04: 94%).

No statistically significant differences were observed between males and females or between classrooms for Teachers A and B indicating that significant differences on change from pre- to posttest were not attributed to gender or teaching styles. Since Teacher A was female and Teacher B was male, it can also be assumed the gender of the teacher did not play a role in changes between pre- and posttest measures.

Correlations calculated on within-program assessments revealed the number of retakes a student took was not correlated to change on indices of percentile rank and stanine, however change on the measure of *SRI* was negatively correlated to the number of retakes a student took. A negative correlation found between number of re-takes and *SRI* change indicates that the more re-takes students took, the less improvement they made on the *SRI* over the course of the study. Negative correlations were also found on each of the other indices of reading function indicating that as number of retakes increased reading level on each measure decreased. These findings are not surprising because those students who did not begin the program with strong reading skills would be expected to need more practice and produce less change throughout the program. This is substantiated by research from Cunningham and Stanovich (1997) where increased exposure to

print is associated with high reading functioning (i.e., The Matthew Effect). However, this finding is not ubiquitous, as overall our analysis showed that at-risk students benefited as much and sometimes more than other learners.

The results of this study are not only statistically significant but educationally significant. The RAND report stated that future research in the area of reading comprehension should be motivated by improving outcomes, "not just for students who are failing in the later grades but for all students who are facing increasing academic challenges" (Snow, 2002, p. xi). Because fourth-grade comprehension skills are essential for further growth in writing and reading skills, research on programs like *Reading Success Level A* are vital to student success in reading (Cunningham & Stanovich, 1998). At the fourth-grade level, text becomes expository and stricter expectations are placed on students in the areas of mathematics, science, and writing (Snow et al., 1998). Although word recognition and fluency are important, as Chall (1983) found, it is not sufficient for students to gain comprehension. The RAND report supports this finding by stating that vocabulary, awareness of language structures, inference and reasoning skills, and cognitive and meta-cognitive strategies when reading are of equal importance to word fluency and crucial to successful reading development (Snow, 2002). In fact, Cunningham and Stanovich (1998) reported that without comprehension reading is meaningless and pointless. In addition, educators should teach these comprehension skills explicitly and systematically (Cunningham & Stanovich, 1997; Snow, 2002). Thousands of studies in the area of reading research support the idea that instruction is best practiced through explicit modeling, guided practice, and independent practice (NICHD, 2000). It is also important that instruction builds upon itself (systematic); teaching the most basic skills first and integrating more difficult steps as mastery is acquired (NICHD, 2000; Snow, 2002). This study revealed that teaching explicit, systematic reading comprehension strategies to fourth graders is likely to increase reading comprehension skills and performance in reading and writing.

Although there are many different kinds of reading curricula in schools today, very few focus on comprehension skills specifically and even fewer still are explicit and systematic in nature. Although phonemic awareness and phonics skills are extremely important and should be taught extensively in the lower grades (K-3), systematic and explicit instruction in comprehension strategies are equally important and probably more vital to future reading success (Chall, 1983; NICHD, 2000; Snow, 2002). The results of this study are consistent with other research such as that of the National Reading Panel Report (NICHD, 2000) and the RAND report (Snow, 2002), emphasizing the need for

systematic, explicit instruction in all areas of reading (basic skills and comprehension skills) in school curricula.

Despite the benefits of this study, some limitations were present. First, the study did not incorporate a true experimental design, hindering experimental control (Whitley, 2002). However, one might argue that the knowledge gained in the natural setting is more beneficial, even when complete experimental control is forfeited. The RAND report states, "improved reading comprehension ability, increased knowledge, and engagement with the text—are of the greatest direct relevance to educators " (Snow, 2002, p. xvi). This "relevance" can only be seen in the natural setting.

Second, although great care was taken in accounting for changes in performance from pre- to posttest, other variables could account for change. Because this research project was conducted to evaluate *Reading Success Level A* for the elementary school and access to a control group was not available, it was not possible to assign students randomly to a control and an experimental group. This study would have been strengthened if we could have compared changes from pre- to posttest to those students in other years who did not receive the *Reading Success Level A* program. Future research should focus on if the same improvements would have resulted using traditional classroom strategies.

Third, this program was implemented in one school that previously had scored high in reading and writing on the WASL. The reading and writing success established by this particular school in subsequent years might hinder generalizability to other populations. Further research should focus on implementing such a program in a failing school to see if similar results are found.

Fourth, formal data were not gathered and analyzed on the generalization of gained skills to content area classes (e.g., mathematics, science). The ability for students to exhibit generalized comprehension strategies to other academic areas assists them in being more efficient at these tasks and, consequently, more successful in school. Thus, future research should determine whether or not these skills gained generalize to content areas in a formal manner.

Finally, a question arises with regard to the maintenance of these skills. It is likely that the level of skill loss is low due to the mastery component of the program; however, no data were gathered measuring the level of skill maintenance. Future research should determine if these skills are maintained across time.

This study showed the successful implementation of the *Reading Success Level A* program with fourth graders. This study provides further data supporting the importance of reading comprehension

skills and the effectiveness of explicit and systematic instruction in the classroom. The goal of educators today should be to implement good instruction because it, "is the most powerful means of promoting the development of proficient comprehenders and preventing reading comprehension problems" (Snow, 2002, p. xvii).

Acknowledgment

References

Armbruster, B. B., Lehr, F., & Osborn, J. (2003). *Put reading first: The research building blocks of reading instruction (Kindergarten through Grade 3)* (2nd ed.). Washington, DC: Partnership for Reading, Center for the Improvement of Early Reading Achievement.

Chall, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill.

Cross, D. R., & Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Developmental Psychology, 80*, 131-142.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Journal of Educational Psychology, 33*, 934-945.

Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator, 22*(1-2), 8-15.

Dixon, B., Klau, K., Rosoff, A., & Conrad, L. (2002). *Reading Success: Level A*. Boston, MA: Classical Learning Universe.

Durkin, D. (1993). *Teaching them to read* (6th ed.). Boston, MA: Allyn and Bacon.

Dymock, S. (1993). Reading but not understanding. *Journal of Reading, 37*(2), 86-91.

Engelmann, S., & Bruner, E.C. (2003). *Reading Mastery Classic.* Columbus, OH: SRA/McGraw-Hill.

Engelmann, S., Johnson, G., Carnine, L., Meyer, L., Becker, W., & Eisele, J. (1998). *Corrective Reading: Decoding B1, B2*. Chicago, IL: SRA/McGraw-Hill.

Fournier, D. N. E., & Graves, M. F. (2002). Scaffolding adolescents' comprehension of short stories. *Journal of Adolescent and Adult Literacy*, *46*, 30-39.

Ihnot, C. (1991). *Read Naturally*. Retrieved on August 31, 2004 from http://www.readnatrually.com.

Lyon, R. G. (1998). Why reading is not a natural process. *Educational Leadership, 55*(6), 14-19.

Lyon, R. G. (1999). *Developing reading skills in young children – lecture summary* (San Francisco, February 20, 1999). Retrieved on June 27, 2004 from http://www.schwablearning.org.

Lyon, R. G., & Moats, L. C. (1997). Critical conceptual and methodological considerations in reading intervention research. *Journal of Learning Disabilities, 30*, 578-589.

Martella, R. C., Nelson, J. R., & Marchand-Martella, N. E. (1999). *Research methods: Learning to become a critical research consumer*. Boston, MA: Pearson Education.

Meese, R. L. (2001). *Teaching learners with mild disabilities: Integrating research and practice* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

National Center for Education Statistics. (2003).The Nation's Report Card. Retrieved on March 17, 2004 from http://www.nces.ed.gov/nationsreportcard/reading.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication NO. 00-4754). Washington, DC: U.S. Government Printing Office.

Osborn, J., Lehr, F., & Hiebert, E. H. (2003). *A focus on fluency. Research-based practices in early reading series by Pacific Resources for Education and Learning*. Retrieved on June 26, 2004 from http://www.prel.org/products/re_/fluency-1.htm.

Rosenshine, B. V. (1987). Explicit teaching and teacher training. *Journal of Teacher Education, 38*(3), 34-36.

Schieffer, C., Marchand-Martella, N., Martella, R., & Simonsen, F. (2002). *The research base for Reading Mastery: Direct Instruction reading*. Desoto, TX: SRA/McGraw-Hill.

Scholastic. (1999). *Scholastic Reading Inventory*. New York, NY: Scholastic

Scholastic. (2002). *Scholastic Reading Counts!*. Retrieved on August 31, 2004 from http://www.src.scholastic.com.

Snow, C. E. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Pittsburgh, PA: RAND.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Retrieved on June 7, 2004 from http://www.nap.edu/catalog/6023.html.

Stenner, A.J., Smith, M., Horabin, I., & Smith, D. (1988). *User's manual for the lexile test of reading comprehension*. Durham, NC: Metametrics.

Tindal, G., Hasbrouck, J., & Jones, C. (2005, January). *Oral reading fluency: 90 years of assessment* (BRT Technical Report No. 33). Eugene, OR: Behavioral Research and Teaching.

Whitely, B.E. (2002). *Principles of research in behavioral science* (2nd ed.). New York, NY: McGraw-Hill.