A Research Study to Investigate The Effects of The *Powerful Vocabulary for Reading Success Program* on Students' Reading Vocabulary and Comprehension Achievement (Research Report 2963-005 of the Institute for Literacy Enhancement) [1]

by

Dr. Cathy Collins Block
Professor of Education
Texas Christian University
(Board of Directors of the International Reading Association 2002-2005)
2800 S. University Drive
Fort Worth, TX 76129
c.block@tcu.edu

and

Dr. John N. Mangieri
Director, Institute for Literacy Enhancement
9807 Hobbitshire Lane
Suite 100
Charlotte, NC 28269
jmangieri@carolina.rr.com

[1] The authors would like to express our gratitude to Ms. Shannon Lewis, Doctoral Candidate in Statistics, for conducting the computer analyses in the data analysis section of this report. We also express our appreciation to Ms. Maggie Wallace, Masters Candidate in Education, for her assistance in data entry for this research project, and Ms. Rachel Escamilla on her Administrative Assistance and typing of the final report.

Vocabulary occupies a contradictory position within America's classrooms. On the one hand, few, if any teacher or literacy professional would argue about the important role that a strong reading vocabulary plays in a student's ability to derive meaning from text. Conversely, over the years, efforts to develop pupils' vocabulary have generally been ineffective and fraught with erroneous instructional practices and learning premises.

This research investigation sought to answer four important questions:

1. If students receive sound, direct, evidenced-based vocabulary instruction for 15 minutes a day, will their reading vocabulary improve significantly?
2. Will these students outperform their counterparts whose instruction in vocabulary is more traditional, limited to the words and strategies contained in literature anthologies?
3. What effects does systematic, high quality, evidenced-based vocabulary instruction have on the vocabulary achievement of third, fourth, and sixth graders who read above, on, or below their grade level placements? and,
4. Can 15 minutes of daily, evidenced-based vocabulary instruction help to close the SES-related achievement gap in reading comprehension?

Theoretical Foundation

In the past two decades a plethora of research investigations have shown the effects of vocabulary instruction on students' literacy growth. Rather than take the space to list and describe each of these well-known investigations, a sampling of some of the more notable ones, which not only shaped the ingredients of the Powerful Vocabulary for

Reading Success program, but also served as the theoretical basis of this investigation, are listed below.

- Beck, Perfetti, and McKeown (1982) demonstrated that 4th graders receiving vocabulary instruction performed better on semantic tasks than those who did not receive instruction.
- McKeown, Beck, Omanson, and Perfetti (1983) also found that vocabulary instruction had a strong relation to text comprehension for 4th grade students.
- Stahl and Fairbanks (1986) conducted a meta-analysis of all vocabulary studies to date and concluded that vocabulary instruction was an important component for comprehension achievement. The best instructional techniques were mixes of definitional and contextual programs; the keyword method produced some significant gains in recall. Repeated exposures to words were also found to be an effective instructional strategy.
- Direct instruction was found to be highly effective for vocabulary learning (Tomeson & Aarnoutse, 1998; White, Graves, & Slater, 1990; Dole, Sloan, & Trathen, 1995; Rinalid, Sells, & McLaughlin, 1997). In addition, the more concrete and personal connections that students can make to a specific word, the better it seems to be learned. For example, there is empirical evidence indicating that making connections with other reading material or oral language in other contexts seems to have large effects in the development of a rich, reading vocabulary.

- It also seems clear from the National Reading Panel's (NICHD, 2001) data set that having students encounter vocabulary words often and in various ways can have a significant effect on the development of increased reading vocabulary (NICHD, 2001; Senechal, 1997; Leung, 1992; Daniels, 1994, 1996; Dole, Sloan, & Trathen, 1995). Although not a surprising finding, this does have direct implications for instruction. Students should not only repeat vocabulary items in learning, but they should be given words that will be likely to appear in many texts and contexts.
- In much the same way that multiple exposures are important, the context in which a word is learned is critical (McKeown, Beck, Omanson, and Pople, 1985; Kameenui, Carnine, & Freschi, 1982; Dole, Sloan, & Trathen, 1995). Vocabulary words should be words that the learner will find useful in many contexts. To that end, a large portion of vocabulary items should be derived from content learning materials.
- The National Reading Panel's (NICHD, 2001) recommendations included the following:

  (a) Vocabulary should be taught both directly and indirectly;

  (b) Repetition and multiple exposures to vocabulary items are important;

  (c) Learning in rich contexts is valuable for vocabulary learning;

  (d) How vocabulary is assessed and evaluated can have differential effects on instruction; and,

  (e) Dependence on a single vocabulary learning strategy will not result in optimal learning.

Method

Participants

This study involved 417 elementary and middle school students who participated in a twenty-two week research study to assess the effects of the Powerful Vocabulary for Reading Success (PVRS) on their reading abilities. This population came from grades 3-6, and of the students in it, 180 of them were in the experimental group. The remaining 237 students were in the control group. Of these children, approximately 30% were above grade-level readers, approximately 27% were on-grade level readers, and the remaining students 43% read below their grade level placements. This population was socioeconomically and ethnically diverse, and represented the spectrum of students who attend schools in the U.S.A.

Students were enrolled in an urban, intercity school; a middle-class suburban school; or, a rural, middle-class, county school. All schools were public institutions of education, had a total school population that was within 23 students of the national mean. Students came from Caucasian (42%), African American (.29%), and Hispanic (19%) origins.

Procedure

The classrooms of student in three elementary schools and one middle school were randomly assigned to experimental and control groups by site coordinators prior to the study. These schools were located in Texas, New Jersey, and Maryland and they were represented the demographics of public institutes in the U.S.A. (as determined by

2000 Census data). In addition to their regular literacy block of instruction, experimental students received approximately 15 minutes of daily instruction (using the Powerful Vocabulary for Reading Success program). As described in greater detail in the Scholastic web site (http://www.teacher.scholastic.com/products/powerfulvocabulary), this program (a) teaches high utility words; (b) provides students with a range of instructional experiences including reading, writing, speaking, and listening; (c) has a detailed teaching plan for each lesson that includes a Think Aloud activity which explicitly models thinking about word meanings; (d) uses metacognitive strategies to aid students in the Acquisition and Retention of words: and, (e) measures a student's mastery of a lesson's objective, which includes vocabulary word meanings, vocabulary building strategies, and word learning principles.

Students in the control group did not receive instruction using the Powerful Vocabulary for Reading Success, but continued to receive their customary literacy instruction for an additional 15 minutes a day. The total block of time devoted to reading instruction in control groups was the same as that which occurred in experimental groups.

Assessment Measures

At the inception of this investigation, all subjects were assessed with the Stanford Nine Achievement Test Battery. This measure was selected because of its quality and congruence with state competency-based literacy tests.

Following the twenty-two weeks of instruction, the aforementioned measures were administered to all subjects as post-tests. The Stanford Nine Achievement Test given at this time was one grade level higher in order to suit the new chronological ages of the subjects. In addition, a metacognitive vocabulary transfer test, which contained

passages from basal readers, not being used in the study's schools, and that had been written between 2001-2005.

When constructing the metacognitive transfer vocabulary test, the researchers completed the following procedures. First, they selected a one-page, single-typed, opening page from a basal reader that had been published since 2001. For example, several of the stories were selected from Macmillan, and Scott Foresman. To be chosen, the stories had to not only appear in the grade level for which they were to be used, but they had to contain a readability of that grade level as measured by the Fry Readability Graph (1977). Next, every passage was read and four words for each category of English vocabulary terms were encircled in a box. Each word was identified as to whether it was a Type 1, 2, 3, or 4 words. Each test contained three Type 1, 2, 3, or 4 words. Tests were constructed so that students would define each of the words that appeared in a box.

This test had directions printed on the top of the page. It was a multiple-choice test. Four possible answers were written for each item. Correct answers were randomly assigned to be choice a, b, c, or d, through the selection procedure of randomly assigning numbers. Answers foils were constructed following these guidelines. Every item had one foil was the definition of a different word that occurred in the same sentence in which the vocabulary word appeared. This foil was not the definition of the word that was in the box. A second incorrect answer choice was the definition of a word that looked similar to the word that was in the box. A third incorrect answer choice was a definition of a word that had a meaning similar to that of the word in the box, but this meaning would be clearly an incorrect answer.

In constructing the test, only one vocabulary word appeared in a sentence, therefore students could use the entire context of that sentence (and those that preceded and followed it) to determine the meaning of each word. There was never an instance when two vocabulary words appeared in the same sentence. None of the words had been taught to students. All experimental and control subjects received a metacognitive vocabulary transfer test that had a readability level which was at their grade level placement and one that was at a readability level which was one grade level above their grade placement.

Researchers independently analyzed each item on the metacognitive transfer test and the Stanford Achievement Subtests to identify the type of vocabulary word or comprehension process being assessed by each item. Researchers' interrater reliabilities ranged from 97% to 85% on these test measures. All disagreements on single items were resolved through a reanalysis and discussion. In addition, a split-half test of reliability was computed for the informal metacognitive vocabulary transfer test. The split-half Pearson correlation coefficient was .93, indicating that the test measured vocabulary reliability. Content validity was established (a) by pattering test items after standardized vocabulary tests, (b) through inter-rater reliability percentages, and (c) by asking for the judgments of forty masters level reading specialists who were asked what domain the test was designed to measure. One hundred percent of these experts judged the test to be a valid assessment of students reading vocabulary.

Data Analysis and Results

In order to examine the effectiveness of the Power Vocabulary for Reading Success program multiple analysis of variance (ANCOVA) was conducted to discern

students' growth in vocabulary, attitude toward reading, and transfer of vocabulary building strategies to novel text. Additionally, data obtained from post-study interviews with experimental group students and teachers were analyzed. In the ANCOVAs, the independent variable, condition, had two levels: experimental and control. The dependent variables were the above-mentioned tests.

The analysis of data was done on a grade-by-grade basis. Within a grade level, vocabulary achievement was measured on the basis of four widely-accepted criteria for words. They were: Type I-words, which frequently appear in students' texts; Type II-words, which contain prefixes, suffixes, and roots; Type III – content specific words and, Type IV – unusual or foreign derivative words. Also within each grade level, comprehension strategies (e.g., drawing conclusion) were assessed in order to determine the impact that this program had upon them. ANCOVAs were adjusted for preexisting pretest effects of gender, ethnicity and pre-study reading ability. Tests of the significant contributing effects of ethnicity or gender were reported when such differences effected post-test score performances.

RESULTS

Teachers were asked to provide the reading level of every student before the study began and after the study ended (as determined by scores on statewide, criterion-based reading tests). At study's end experimental students' overall mean reading ability rose significantly above the control subjects' mean [$F(1,32) = 32.71$, $p < .001$.]

Shown below are the results of ANCOVAs for each grade level.

Third Grade

The mean scores on the metacognitive vocabulary transfer test of Type I words, which were adjusted for initial ability differences, showed that third-grade experimental subjects significantly outperformed control students on the post-test of the Stanford Vocabulary Subtest portion which measured Type I words ($F(1,29) = 9.585$, $p = .004$). Experimental group means on Type I words was 6.94 (S.D.=1.06); control group mean was 5.44 (S.D.=1.5).

The mean scores on the same vocabulary transfer assessment for Type II words, which were adjusted for initial ability differences, showed that third-grade subjects experimental students significantly outperformed the control students on the post-test of the Stanford Vocabulary Subtest portion which measured Type II words ($F(1,29) = 9.670$, $p = .03$). Experimental group means on Type II words was 7.69 (S.D.=1.25); control group mean was 6.56 (S.D.=1.37).

The mean scores on the same vocabulary transfer assessment for Type III words, which were adjusted for initial ability differences, showed that experimental third-graders significantly outperformed control group students' on the post-test of the Stanford Vocabulary Subtest portion which measured Type III words ($F(1,29) = 4.874$, $p = .04$). Experimental mean was 8.06 (S.D.=1.06); control mean was 6.94 (S.D.=1.65).

The mean scores on the same vocabulary transfer test for Type IV words, which were adjusted for initial ability differences, showed that the experimental group outperformed the control group on the post-test of the Stanford Vocabulary Subtest portion which measured Type IV words ($F(1,29) = .280$, $p = .601$, N.S.) Experimental mean was 3.0 (S.D. =1.16); control mean was 2.63 (S.D. =1.26). The non-significant

effects on Type IV words was expected because no student in the experimental group was taught strategies to learn Type IV words. The lessons that taught these strategies in Powerful Vocabulary for Reading Success had not yet been taught by experimental teachers as they were the last 8 lessons in the program and post-tests were given in March/April of the school year, prior to this instruction.

However, the total mean scores on the post-Stanford Total Vocabulary Subtest, which were adjusted for initial ability differences, showed that the experimental group significantly outperformed the control group in total subtest scores ($F(1,29) = 9.376$, $p < .05$). Experimental mean was 25.69 (S.D.=2.89); control mean was 21.56 (S.D.=4.65).

In no area of comprehension did the control group students outperform the experimental group students. In addition, on the Stanford Comprehension Subtest, experimental groups significantly outperformed control groups in the areas of: (a) main idea [experimental mean =4.69 (S.D.=.95); control mean =3.94 (S.D.=1.77), $F(1,29) = 7.91$, $p = .009$]; remembering details [experimental mean = 14.31 (S.D. =2.3); control mean =13.13 (S.D.=3.3), $F(1,29) = 8.07$, $p = .008$]; (c) inference [experimental mean =8.56 (S.D.=1.5); control mean =8.25 (S.D.=1.84), $F(1,29) = 7.22$, $p = .012$]; and (d) total Comprehension Subtest mean scores [experimental mean =39.6 (S.D.=5.66); control mean =37.31 (S.D.=9.0), $F(1,29) = 11.15$, $p = .0001$, $n^2 = .40$].

Experimental group subjects outperformed control group subjects on the sequence items of the post-test of the Stanford Comprehension Subtest [experimental mean = 1.75 (S.D. =.45); control mean = 1.62 (S.D. =.89)], and, on drawing conclusion items [experimental mean = 5.6 (S.D. =1.36); control mean = 5.69 (S.D. =1.58)], but

when these differences were adjusted for initial pretest abilities, means were not significantly different, p=.30 and p =.97, respectively.

When experimental and control mean scores on the Metacognitive Vocabulary Transfer Test (ability to determine the meaning of Type I words that had never been taught in a passage which has not been read during the treatment period) were analyzed, experimental subjects significantly outperformed control subjects [experimental mean =98.3 (S.D. =6.45); control mean =72.2 (S.D. =30.78), F (1,31) =10.36, p =. 003]. Experimental subjects' mean scores were also higher than control subjects' mean scores on the Metacognitive Vocabulary Transfer Test of Type II; III, and Type IV words and these means approached significance [Type II experimental mean =86.7 (S.D. =18.6); control mean =75.0 (S.D. =29.7); Type III experimental = 76.6 (S.D. =19.97); control mean =65.28 (S.D.=24.5); Type IV words experimental mean = 88.33 (S.D. =15.99); control mean = 28.57 (S.D. =18].

When total post-test scores on the Stanford Vocabulary Subtest were controlled for gender, ethnicity, and pre-test scores on the Stanford Vocabulary Subtest, the interaction was significant, with experimental groups outperforming control groups matched by ethnicity and gender [F (1,27) = 8.55, p=.007, n^2 =.24]. African American control subjects' mean score fell from 24 (S.D. =3.69) to 22.57 (S.D. =4.78); Hispanic control subjects' mean score fell from 23.36 (S.D. =5.08) to 21.13 (S.D. =5.03); and, Caucasian control subjects' mean score fell from 22.5 (S.D. =2.36) to 18.0 (S.D. =6.36) in the 22 weeks that the vocabulary study occurred. By contrast, African American, Hispanic, and Caucasian experimental subjects' mean scores rose significantly during the

treatment period. Their mean scores were 25.11 (S.D. =2.5); 26.33 (S.D. =3.6); and 27 (S.D. =6.36) respectively.

Fourth Grade

The mean scores on the vocabulary transfer test of Type I words, which were adjusted for initial ability differences, showed that fourth-grade experimental groups significantly outperformed fourth grade control groups on the post-test of the Stanford Vocabulary Subtest portion which measured Type I words, ($F(1,23) = 6.669$, $p = .017$), $n^2 = .23$). Experimental group means on Type I words was 3.07 (S.D. =1.49); control group mean was 1.93 (S.D. = .92).

The mean scores on the vocabulary transfer test of Type II words, which were adjusted for initial ability differences, showed that fourth-grade experimental groups significantly outperformed fourth-grade control groups on the post-test of the Stanford Vocabulary Subtest portion which measured Type II words, ($F(1,25) = 4.932$, $p = .04$, $n^2 = .36$ for condition. Experimental group means on Type II words was 2.86 (S.D. = 1.35); control group mean was 1.79 (S.D. =.98).

The mean scores on the vocabulary transfer test of Type III words, which were adjusted for initial ability differences, showed that the experimental group significantly outperformed the control group on the post-test of the Stanford Vocabulary Subtest portion which measured Type III words ($F(1,23) = 4.48$, $p = .05$). Experimental mean was 3.79 (S.D. = 1.67); control mean was 3.0 (S.D. = 1.24, with $n^2 = .12$ for condition.

The mean scores on the vocabulary Type IV transfer assessment, which were adjusted for initial ability differences showed that the experimental group outperformed the control group on the post Stanford Vocabulary Subtest portion which measured Type

IV words (F (1,20) =2.97, p =.07, n^2 = .23 for condition). Experimental mean was 3.09 (S.D. = 1.16); control mean was 2.63 (S.D. =1.26). These data suggest that experimental subjects independently transferred vocabulary building strategies that were learned from Type I – III lessons to Type IV words. Type IV vocabulary building strategies and word learning principals were not yet taught to these fourth graders, prior to the post-test measures.

The mean total scores on the post Stanford Vocabulary Subtest, which were adjusted for initial ability differences, showed that the experimental group significantly outperformed the control group (F (1,23) = 5.211, p<.03). Experimental means was 14.79 (S.D. =6.65); control mean was 9.57 (S.D. =3.36), n^2 =.31 for condition.

In no area of comprehension did control fourth-graders outperform experimental fourth-graders on the Stanford Comprehension Subtest. Specifically, the experimental group students outperformed the control group students in the areas of : (a) main idea [experimental mean =1.71 (S.D. =1.66); control mean =.50 (S.D. =.65); F (1,23) = 4.66, p = .04; n^2 = .20 for condition]; (b) remembering details [experimental mean =4.86 (S.D. =3.5); control mean = 3.93 (S.D. =.14); F (1,23) = .528, p =.48]; (c) inference [experimental mean =7.6 (S.D. =4.3); control mean = 5.71 (S.D. =2.3) F (1,25) = 7.20, p = .013) n^2 = .22 for condition], and (d) metacomprehension mean scores [experimental mean = 1.36 (S.D. = .88); control mean = .36 (S.D. =.63), F (1,24) = 6.634, p =.02, =n^2 =.22 for condition]. Similarly, total Stanford Comprehension Subtest mean scores for experimental subjects were significant higher than control subjects' mean scores, (F (1,23) =5.05, p =.04, n^2 =.40. Experimental mean was 21.93 (S.D =11.97). Control mean was 14.29 (S.D. = 4.99).

Experimental subjects outperformed control subjects on the sequence items of the post the Stanford Comprehension Subtest [experimental mean =.71 (S.D. =.36); control mean = .36 (S.D. = .50)], and on drawing conclusion items [experimental mean =5.6 (S.D. =2.90); control mean =3.43 (S.D. =2.10)], but these adjusted for initial ability means were not significantly different, when measuring sequence and drawing conclusions skills, p=.19 and p=.21, respectively.

At the fourth-grade level, Powerful Vocabulary Reading Success program demonstrated to significantly increase male experimental subjects abilities to remember details on the post-test Stanford Comprehension Subtest [F (1,23) =5.54, p= .03, n^2 =.19 for gender effects]. Female means were identical on this posttest for control and experimental subjects at 3.0 (S.D. for experimental =1.10; and S.D. for control subjects was 3.34). Male experimental mean was 5.89 (S.D. = 1.8) and control mean was 4.63 (S.D. =1.7).

Fifth Grade

The mean scores on the vocabulary transfer test of Type I words, which was adjusted for initial ability differences, showed no significant differences between the experimental and control group on the post Stanford Vocabulary Subtest portion which measured Type I words. Experimental group means on Type I words was 6.23 (S.D. =2.28); control group mean was 7.0 (S.D. =2.18). These data are not surprising as the experimental teacher did not teach all lessons for Type l words.

The mean scores on the vocabulary transfer test of Type II words, which were adjusted for initial ability differences, showed that the experimental group of fifth-graders significantly outperformed the control group of fifth-graders on the post-test of the

Stanford Vocabulary Subtest portion which measured Type II words, [F (1,27) =7.416, p =.01, n^2 =.22 for condition effect]. Experimental group means on Type II words was 4.35 (S.D. =2.0); control group mean was 3.77 (S.D. =1.64).

The mean scores on the vocabulary test of Type III words, which were adjusted for initial ability differences, showed that at the fifth-grade level the experimental and control groups did not differ significantly on the post-test of the Stanford Vocabulary Subtest portion which measured Type IIII words. Experimental mean was 3.23 (S.D. =1.24); control mean was 3.71 (S.D. =1.93). Similarly, the mean scores on the vocabulary Type IV assessment, which were adjusted for initial differences showed that third-grade experimental and control groups did not differ significantly on the post Stanford Vocabulary Subtest portion which measured Type IV words. Experimental mean was 5.23 (S.D. =1.55); control mean was 5.53 (S.D. =1.55). These results were expected as experimental fourth-graders did not receive instruction I these strategies. The total mean scores on Stanford Vocabulary Subtest, which were adjusted for initial differences, showed that experimental and control fifth-graders did not differ significantly. Experimental mean was 18.46 (S.D. =5.49); control mean was 20.59 (S.D. =6.47). These results were expected as a review of individual student books revealed that no experimental fifth grader received instruction in vocabulary building strategies for Type III and IV words, and only a few lessons in Type I words.

Sixth Grade

The means scores on the vocabulary transfer test of Type I words, which was adjusted for initial ability differences, showed that experimental sixth-graders significantly outperformed control groups on the post Stanford Vocabulary Subtest

portion which measured Type I words, [F (1,130) = 7.495, (p=.001, n^2 =.1 effect for condition]. Experimental group means on Type I words was 7.7 (S.D. =1.7); control group mean was 7.22 (S.D. =2.182).

The mean scores on the vocabulary transfer test of Type II words, which was adjusted for initial ability differences, showed that experimental sixth-graders outperformed control groups on the post Stanford Vocabulary Subtest portion which measured Type II words. Experimental group mean was 6.84 (S.D. =1.10); control group mean was 6.35 (S.D. =2.02). Similarly, the mean scores on the vocabulary transfer test of Type III words, which was adjusted for initial ability differences, showed that experimental groups outperformed control groups on the post Stanford Vocabulary Subtest portion which measured Type III words. Experimental mean was 3.75 (S.D. = 1.48); control mean was 3.41 (S.D. = .45).

On the other hand, the mean scores on the vocabulary transfer of Type IV words, which was adjusted for initial ability differences, showed the sixth-grade experimental groups significantly outperformed the control groups on the post Stanford Vocabulary Subtest portion which measured Type IV words, [F (1,130) =8.91), p =.003]. Experimental mean was 1.56 (S.D. =.85); control mean was 1.13 (S.D. =.81). The partial n^2 = .064 suggests a small and significant relationship between the treatment and students' scores on the post-treatment vocabulary Type IV assessment, controlling for pre-treatment vocabulary Type IV proficiency. Condition accounted for 6.4% of the variance of the dependent variable, holding constant the per-test scores on the vocabulary IV questions. Sixth-graders in the experimental group scores rose significantly higher on the Type IV vocabulary words at the end of the study (M =1.56) when compared to the

control-group sixth graders (M =1.13), when controlling for initial scores on the Type IV vocabulary assessment. A subsequent analysis of covariance was computed controlling for ethnicity and gender in addition to pretest scores. The ANCOVA remained significant, $F(1,128) = 9.077$, $p = .033$, $n^2 = .066$. In like manner, the mean scores on post Stanford Total Vocabulary Subtest, which were adjusted for initial differences, showed that experimental sixth-graders significantly outperformed their control group peers ($F(1,132) = 4.12$, $p < .03$). Experimental mean was 24 (S.D. =4.55); control mean was 21.86 (S.D. =6.02).

In no area of comprehension did sixth-grade control group students outperform the experimental group sixth-graders. On the Stanford Comprehension Subtest, experimental students outperformed the control subjects in the areas of: (a) main idea [experimental mean =7.76 (S.D. =2.76); control mean = 7.41 (S.D. =3.22)], remembering details [experimental mean = 9.48 (S.D. =3.5); control mean =8.76 (S.D. =3.02)]; (c) inference [experimental mean =8.55 (S.D. =2.8); control mean = 8.33 (S.D. =3.17)], and (d) total Comprehension Subtest mean scores [experimental mean = 33.89 (S.D. =12.59); control mean =33.51 (S.D. = 9.88)].

Experimental group subjects outperformed control group subjects on the metacomprehension items of the post-test the Stanford Comprehension Subtest [experimental mean = 4.0 (S.D. =1.78); control mean = 3.74 (S.D. =1.65], and on drawing conclusion items [experimental mean =4.75 (S.D. =1.77); control mean = 4.67 (S.D. = 2.02)].

When experimental and control mean scores on the Metacognitive Vocabulary Transfer Test (ability to determine the meaning of Type I words that had never been

taught in a novel passage that had never been read during the treatment period), above-grade-level, experimental subjects significantly outperformed above-grade-level, control subjects [experimental mean = 88.3(S.D. =23.43); control mean = 70.2 (S.D. =28.8, $F(1,68) = 6.196$, $p = .013$, $n^2 = .08$ for condition]. Sixth-grade experimental subjects' mean scores were also significantly higher than sixth-grade control subjects mean scores on Type II, III, and Type IV words on the Metacognitive Vocabulary Transfer Test for above-grade level readers. For Type II words, above-grade level sixth-grade readers' experimental mean score was 81.66 (S.D. =20.69) and above-grade level control readers' mean score was 60.12 (S.D. =27.64), $F(1,68) = 11.948$, $p = .007$, $n^2 = .15$ for condition. For Type III words, above-grade level sixth-grade experimental students' mean score was 70.83 (S.D. =29.4); control mean was 49.40 (S.D. =35.6), $F(1,68) = 5.502$, $p = .022$, $n^2 = .08$ for condition. For Type IV words, above grade-level sixth-grade experimental mean score was 86.67 (S.D. =15.72); control mean was 71.42 (S.D. =31), $F(1,68) = 4.815$, $p = .032$, $n^2 = .07$ for condition.

Conclusions

In the numerous research investigations, which have been conducted in the area of vocabulary, various components in the Powerful Vocabulary for Reading Success program have been assessed as to their importance relative to students' abilities to learn the meaning of words. For example, direct instruction has been consistently found to be an effective strategy for vocabulary learning. This investigation was unique from prior studies, however, in that it assessed the effect of not a single measure but rather the usage of (and integration of) five components in a single program's effects on vocabulary and comprehension achievement. These components were described in Procedure section of

this document. The study's data provide the literacy community with valuable information relative to vocabulary instruction. These data led to the following conclusions.

1. Rather than vocabulary being only a small a part of a larger literacy lesson (as occurred in control group classrooms), when vocabulary (which contains the learning principles of the Powerful Vocabulary Reading for Success program), is directly taught for 15 minutes a day, students in grades 3-6 will learn the meaning of a significantly greater numbers of words than peers whose instruction is not as research-based, systematic, and direct.
2. A vocabulary program that teaches all four Types of words and does so with a unifying theme (e.g., only multiple meaning words are taught in one lesson while only prefixes are taught in a subsequent lesson), as is found in the Powerful Vocabulary Reading for Success program, demonstrated to produce significantly greater vocabulary growth in each of these four classifications of English word types. This finding was true for students in grades 3-6.
3. When confronted with unfamiliar words in unfamiliar texts, the experimental groups determined the meaning of a significantly greater number of words than students in the control group who did not receive instruction from the Powerful Vocabulary for Reading Success Program.
4. The linkage between vocabulary growth and reading comprehension success is a strong one. In each instance, after students participated in the Powerful Vocabulary for Reading Success program for twenty-two weeks, vocabularies

enabled them to comprehend text significantly better than students in control group in grades 3-6.

5. These gains in overall comprehension (see #3 above) also translated into increased prowess in both literal and implicit comprehension subtest areas. For example, experimental group students attained appreciably higher scores on both the Total Stanford Comprehension Subtest as well as on the transfer test in subskill items of finding main ideas, remembering details, and inferencing.
6. While it was noted that experimental subjects learned significantly more Type I, II, III, and IV words than control students, the Powerful Vocabulary for Reading Success program, in less than a year's time, produced especially large gain in Type I words. This finding is important in that Type I words are words which frequently appear in students' texts, and were the strategies taught in every experimental class (in some of the experimental classes Type II, III, and IV word lessons were not taught). As shown in the comprehension findings of this investigation (see #4 and #5 on the previous page), this newfound ability manifests itself into these students' increased comprehension proficiencies.
7. By grade 3, the adverse effects of SES-related literacy achievement gaps are well-established within students. The instructional facets of the Powerful Vocabulary for Reading Success program not only counterbalance these negative conditions, but they also eradicate them in that lower SES students made comparable gains in vocabulary and comprehension to those of their more advantaged peers.

8. Teaching students to infer is a key facet of the Powerful Vocabulary for Reading Success program. This instructional emphasis created within experimental group students a metacognitive awareness of how to infer the meaning of not only Type I, II, III, and IV words but also to derive inferential meanings, as they comprehended text.
9. At the end of their twenty-two week's participation in Powerful Vocabulary for Reading Success, experimental students who were functioning below-grade level in week one of their study, grew as rapidly in their vocabulary and comprehension abilities as peers in their class who were performing at, or above, grade-level.
10. The need for intense vocabulary instruction extends into middle school. Even above-grade level, sixth-graders who received daily, direct vocabulary instruction significantly out-performed ability-matched control subjects (on informal and standardized tests of vocabulary and comprehension).

Reference

Beck, I.L., Perfetti, C.A., & McKeown, M.G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. Journal of Educational Psychology, 74(4), 506-521.

Daniels, M. (1994). The effect of sign language on hearing children's language development. Communication Education, 43(4), 291-298.

Dole, J.A., Sloan, C., & Trathen, W. (1995). Teaching vocabulary within the context of literature. Journal of Reading, 38(6), 452-460.

Kameenui, E., Carnine, D., & Freschi, R., (1982). Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. Reading Research Quarterly, 17(3), 367-388.

Leung, C.B. (1992). Effects of word-related variables on vocabulary growth repeated read-aloud events. In C.K.Kinzer & D.J.Leu (Eds.), Literacy research, theory, and practice: Views from many perspectives: Forty-first Yearbook of the National Reading Conference (pp.491-498). Chicago, IL: The National Reading Conference.

McKeown, M.G., Beck, I.L., Omanson, R.C., & Perfetti, C.A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. Journal of Reading Behavior, 15(1), 3-18.

McKeown, M.G., Beck I.L., Omanson, R.C., & Pople, M.T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. Reading Research Quarterly, 20(5), 522-535.

NICHD (2001). Report of the National Reading Panel. National Institute for Child Health and Development. Washington, DC: NICHD.

Rinaldi, L., Sells, D., & McLaughlin, T.F. (1997). The effects of reading racetracks on the sight word acquisition and fluency of elementary students. Journal of Behavioral Education, 7(2), 219-233.

Senechal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. Journal of Child Language, 24(1), 123-138.

Stahl, S.A., & Fairbanks, M.M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. Review of Educational Research, 56(1), 72-110.

Tomesen, M., & Aarnoutse, C. (1998). Effects of an instructional programme for deriving word meanings. Educational Studies, 24(1), 107-128.

White, T.G., Graves, M.F., & Slater, W.H. (1990). Growth of reading vocabulary in diverse elementary schools: Decoding and word meaning. Journal of educational Psychology, 82(2), 281-290.