Educational Research
Institute of America
www.eriaonline.com • Email info@eriaonline.com
An affiliate of Beck Evaluation & Testing Associates, Inc.

Web Version

Technical Report Number 120

A Study of the Instructional Effectiveness of *Power Up!™ Building Reading Strength*

July 2003

A Study of the Instructional Effectiveness of *Power Up!™ Building Reading Strength*

Introduction

This report describes a study of the instructional effectiveness of *Power Up!™ Building Reading Strength*, a supplemental reading program created by Steck-Vaugn Company for middle school students reading several grades below level (2.5–5.5). The program spans four levels and is comprised of paperback books, student workbooks, teacher editions, teacher resource binders with reproducible worksheets, software and web-based components.

Background Information

Steck-Vaughn Company, a division of Harcourt Supplemental Publishers based in Austin, Texas, is a publisher of supplementary educational materials for students in grades K-12 and adult learners. Harcourt Supplemental Publishers contracted with the Educational Research Institute of America (ERIA) to conduct an instructional validation study to evaluate the effectiveness of the new *Power Up!* program. Harcourt Supplemental Publishers supplied schools with all of the instructional materials. ERIA conducted all phases of the study.

Purpose of the Study

Recent statistics concerning reading at the middle grades have given educators much alarm. Though the 2002 National Assessment of Educational Progress (NAEP) results show that eighth graders are reading better than they were in 1998, the improvement is slight—and still only 33% of students in Grade 8 are reading at or above a proficient level. As students enter the middle grades they are expected to comprehend more complex texts, and yet many of these students cannot successfully process grade-level material.

The *Power Up!* program was developed to meet the needs of students at the middle school level who were having significant difficulty in reading. The purpose of this study was to evaluate how effective the program is at boosting reading comprehension for these students, as evidenced through standardized test scores.

Research Questions

The following research questions guided the design of the study:

- Is the *Power Up!: Building Reading Strength* program instructionally effective?
- Do students' reading comprehension skills improve as a result of using the program?

Design and Procedures of the Study

Study Population

The study was conducted with students in grades 6–8 at nine different middle schools in East Providence, Rhode Island; Las Vegas, Nevada; Los Angeles, California; and Terre Haute, Indiana. The schools were all located within urban school districts with an average minority population of 42%. Poverty rates across the districts averaged 17%.

The classes chosen for inclusion in the study were comprised of students experiencing significant difficulty in reading. Most of the students participating in the study were described by their teachers as reading two or more years below grade level. No additional specific test evidence was available to validate the teachers' judgments. However, the pretest scores used in this study showed that the students were reading below grade level. An examination of the pretest scores for the 6th grade students shows pretest scores that are slightly below grade level; the 7th grade students were far below grade level; and the 8th grade students' scores were below grade level but not as much as the 7th grade students. Students' reading levels at the start of study (as indicated by Total Reading Scores on pretests) were as follows:

Sample	Grade Equivalent Reading Level
Grade 6	5.6
Grade 7	3.9
Grade 8	5.0

In total, 208 students across the three grade levels remained in the sample throughout the study. The 34 teachers in this study were volunteers using the *Power Up!* program.

Research Design

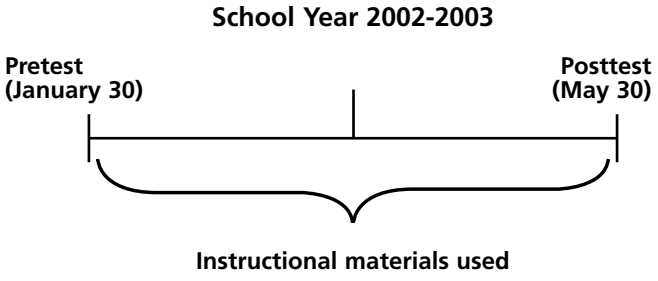
A quasi-experimental pretest/posttest design was used. At the start of the study, students were administered a pretest to evaluate their vocabulary and comprehension skills. Following approximately one semester of instruction using the *Power Up!* materials, students took a posttest in the same skill areas.

The average scores for Vocabulary, Comprehension, and Total Reading were compared using a paired comparison *t*-test. The standard score for each of these three test scores was selected for statistical analysis since it has more stable statistical properties than raw scores.

In addition to the *t*-test analyses and effect size calculations, graphs showing achievement gains or losses were developed using percentile and grade equivalent scores. These were based on the *Stanford Achievement Test*, 9th Edition (SAT 9) national norm sample. The norm scores were determined from the standard scores utilizing tables supplied by the publisher.

Study Timeline

The study was conducted during the second semester of the 2002-2003 school year. It spanned approximately four months. The timeline for the study follows:



Instructional Materials

Harcourt Supplemental Publishers provided teachers with the full set of *Power Up!: Building Reading Strength* program components. Components included paperback books, student workbooks, teacher editions of the student workbooks, teacher resource binders with reproducible worksheets and assessments, CD-ROMs containing dynamic versions of the paperback book texts, and access to the program's web components. Harcourt Supplemental Publishers also provided on-site training for teachers to help them best understand and use the program. Teachers were allowed to use the program at their own discretion. It is not know how extensively each teacher used each component, but we do know that instruction using the *Power Up!* program centered around the paperback books, the student workbooks, and teacher editions.

Outcome Assessments

ERIA selected the abbreviated version of the *Stanford Achievement Test*, 9th Edition (SAT 9) for the pre and posttests. Because of the students' low reading levels, the *Stanford Achievement Test*, *Primary 3* was chosen for use in this study. This level of the SAT 9 is designed for average students enrolled in Grade 4. The SAT 9: Abbreviated, Primary 3 includes two reading tests, Reading Vocabulary and Comprehension. A score is generated for each test as well as for Total Reading, the sum of both.

Results of the Analyses

A separate analysis was conducted for each grade level. Standard scores were used for the statistical analyses. Percentiles and grade equivalent scores were used to compare the students to the norm-referenced population for the SAT 9.

A paired comparison *t*-test was computed to determine the statistical significance of any gain scores. The standard score was used for all of the analyses. The standard scores provide a common metric across all levels of the test. They are much more appropriate for conducting statistical analyses since they are developed to approximate a normal distribution and raw scores are often not normally distributed. All of the norm scores—including percentiles, stanines, and grade equivalents—are derived from the standard scores. These other norm-referenced scores do not have normal distributions and nor do they approximate an interval scale as do the standard scores.

In addition to the *t*-test analyses, in order to determine the size of the effect of the use of the *Power Up!* program, Cohen's *d* was computed. Cohen's *d* is an indication of the strength of the effect. Cohen has provided a general interpretation of the strength of the effect statistic. He suggests that an *r* between 0 and .2 indicates a small effect; an *r* between .3 and .8 indicates a moderate effect; and an *r* of .8 and above indicates a large effect.

Grade 6 Results: *Stanford Achievement Test, Primary 3*

For Grade 6 students the average standard scores increased from pretest to posttest for Reading Vocabulary, Reading Comprehension, and Total Reading. However, the increases were not statistically significant for Reading Vocabulary as can be seen in Table 1. The increases in Reading Comprehension and Total Reading were significant beyond the .01 level. This means that such changes in scores would be expected to happen by chance fewer than once out of a hundred times.

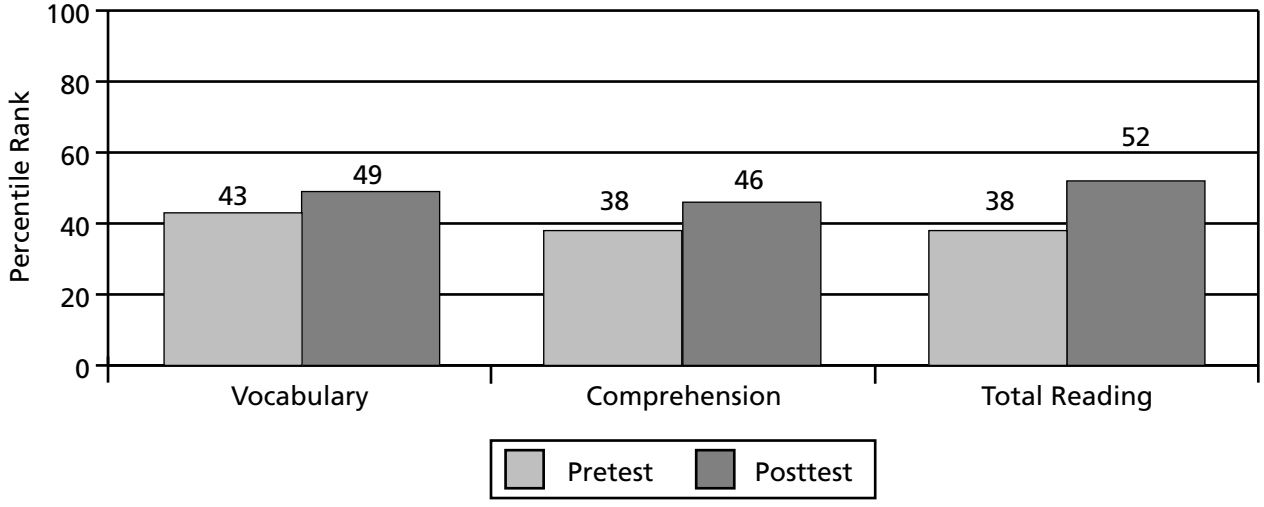
The lack of a statistically significant increase for Reading Vocabulary may have been because the Grade 6 students scored at relatively high levels on both the pretest and the posttest. To provide some idea as to how high those scores were, it is useful to examine the average raw scores for the students on the 20-item Vocabulary test. The average raw score average was 17.12 on the pretest and 17.65 on the posttest. A moderate effect size was noted for the Comprehension and the Total Reading Score; while the magnitude of effect size *r* for Vocabulary was considered to be small.

Table 1
Paired Comparison *t*-test Gains from Pretest to Posttest for Grade 6 Students on the SAT 9, Abbreviated
N=63

Test Score	Pretest Standard Scores		Posttest Standard Scores		Paired Comparison *t*-test		Effect Size	
	Mean	*Standard Deviation*	*Mean*	*Standard Deviation*	*t-test*	*Significance*	*Cohen's d*	*Effect Size r*
Vocabulary	657.55	41.85	663.53	33.28	1.477	<.145	.152	.075
Comprehension	646.98	64.02	660.87	57.47	3.108	<.01	.228	.113
Total	650.00	52.49	661.34	45.88	3.072	<.01	.230	.114

The percentile comparisons of the pretest to posttest gains are shown in Figure 1. These percentile comparisons are based on Grade 6 norms, as these students were enrolled in sixth grade at the time of the study. The SAT 9 provides norm tables that describe how Grade 6 students performed nationally on the Primary 3 level of the SAT 9 Abbreviated.

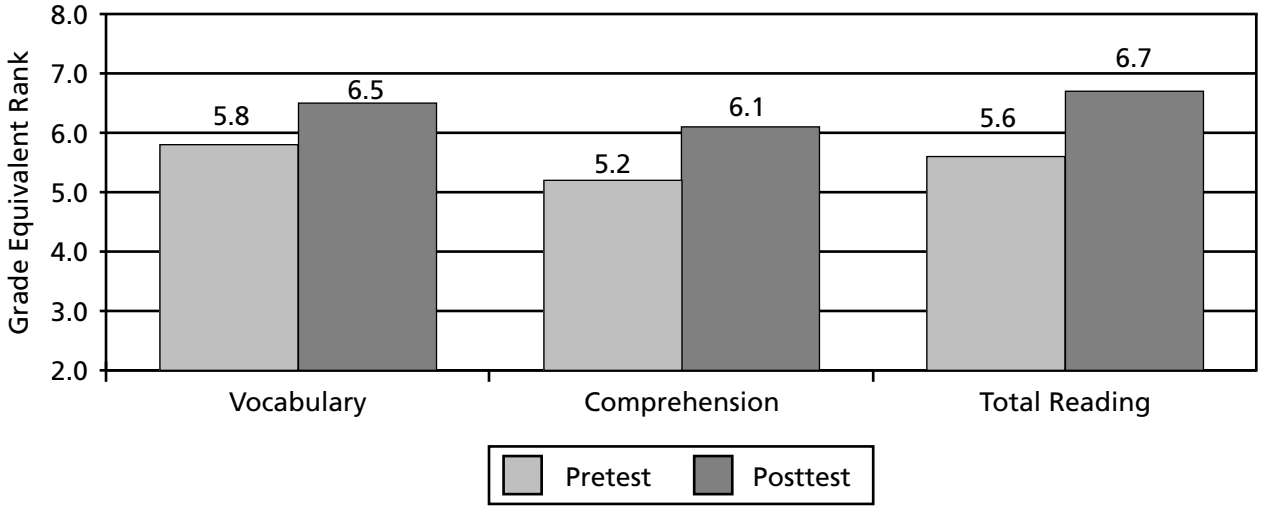
Figure 1
Comparison of Pretest/Posttest Percentile Scores on the SAT 9 Using Grade 6 National Norm Comparisons



The SAT 9 also provides grade equivalent scores to interpret the test scores on a norm-referenced basis. The gains in grade equivalent scores for the Grade 6 students are shown in Figure 2. It appears that these sixth grade students were not reading a great deal below grade level at the beginning of the study despite teachers' assertions that the students were reading well below grade level. This may account for the reason that they achieved such high scores on the Vocabulary subtest. The Primary 3 level Vocabulary test of the SAT 9 may have been a test that was at too low a level for these students. However, the comprehension scores indicate that the test was at an appropriate level. Since the *Power Up!* program focuses primarily on reading comprehension, the test level seems to have been an appropriate choice overall.

The comparison of the grade equivalent gains to the national averages compares gains made by these students to the gains made by the national norm population over a similar time period. Since the study took place over a 4-month period, the gain scores for an average group of students may have been expected to be about .4; however, the total reading grade equivalent gains for these students, who presumably had never made average gains in their previous years of education, averaged .9 across the three scores.

Figure 2
Comparison of Pretest/Posttest Grade Equivalent Scores on the SAT 9 Using Grade 6 National Norm Comparisons



Grade 7 Results: Stanford Achievement Test, Primary 3

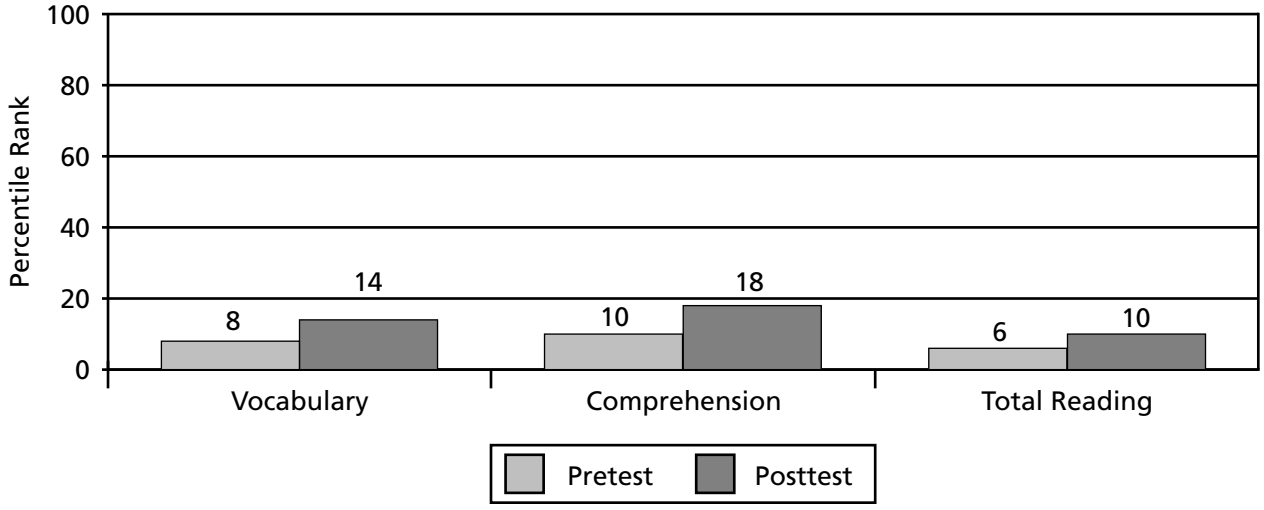
For Grade 7 students, the average standard scores increased from pretest to posttest for Reading Vocabulary, Reading Comprehension, and Total Reading. Increases were statistically significant for all three scores, as can be seen in Table 2. The increases in Reading Comprehension and Total Reading were significant beyond the .0001 level, while the Reading Vocabulary increases were significant at the .001 level. This means that such increases in Reading Comprehension and Total Reading would be expected to happen by chance fewer than once out of ten thousand times, while the increases in Reading Vocabulary would be expected to occur by chance fewer than once out of a thousand times. Although pre/posttest scores moved significantly in a positive direction, given the small effect size for all three scores, the magnitude of the result is not very strong.

Table 2
Paired Comparison *t*-test Gains from Pretest to Posttest
for Grade 7 Students on the SAT 9, Abbreviated
N=92

Test Score	Pretest Standard Scores		Posttest Standard Scores		Paired Comparison *t*-test		Effect Size	
	Mean	*Standard Deviation*	*Mean*	*Standard Deviation*	*t-test*	*Significance*	*Cohen's d*	*Effect Size r*
Vocabulary	624.47	40.81	635.84	43.97	3.293	<.001	.268	.132
Comprehension	616.58	46.55	631.57	45.28	4.976	<.0001	.326	.161
Total	618.14	37.28	631.34	38.10	6.269	<.0001	.350	.172

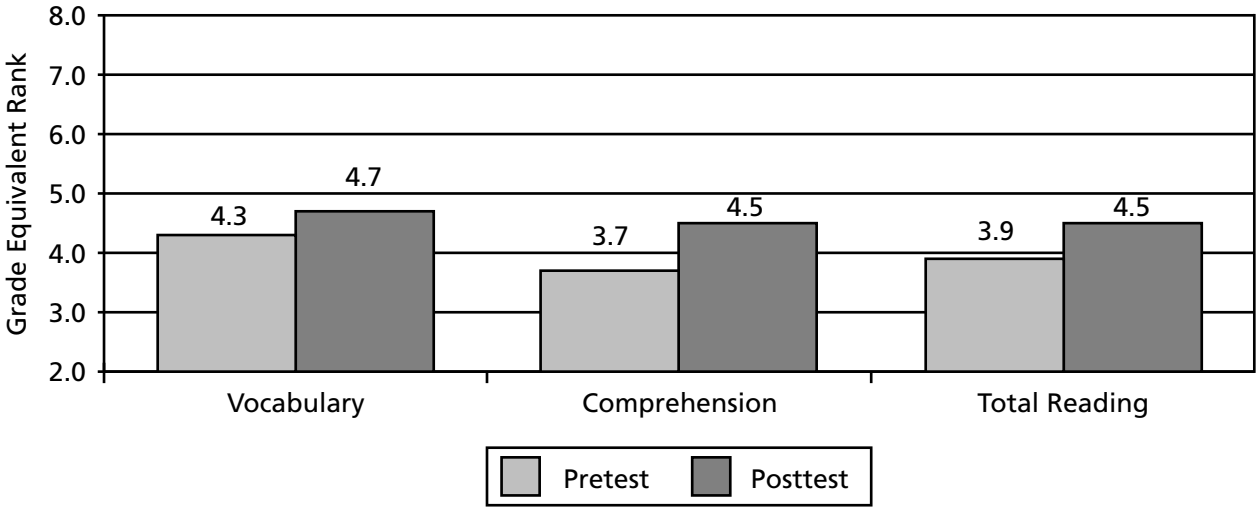
The percentile comparisons of the pretest to posttest gains are shown in Figure 3. These percentile comparisons are based on Grade 7 norms, as these students were enrolled in seventh grade at the time of the study. The SAT 9 provides norm tables that describe how Grade 7 students performed nationally on the Primary 3 level of the SAT 9 Abbreviated. The norm tables give norms for different grade groups on each level of the test. The Grade 7 students were reading at a lower level compared to their norm group than were the students in the Grade 6 sample when they were compared to their norm group. At the beginning of the study, Grade 7 students were only at about the 10th percentile level when compared to the Grade 7 students in the national norm population. Nevertheless, they did show significant gains by the end of the study, as shown in Table 2. They also made gains in terms of percentile increases, averaging an 8 percentile point increase across the three scores.

Figure 3
Comparison of Pretest/Posttest Percentile Scores on the SAT 9,
Abbreviated Test Using Grade 7 National Norm Comparisons



The SAT 9 also provides grade equivalent scores to interpret the test on a norm-referenced basis. The gains in grade equivalent scores for the Grade 7 students are shown in Figure 4. Figure 4 provides further evidence of the low reading levels of the Grade 7 students at the beginning of the study. It appears that these seventh grade students were reading more than three years below grade level at the beginning of the study. The comparison of these grade equivalent gains to the national averages is, however, greater than the gains made by the average students in the national norm group. Since the study took place over a 4-month period, the gain scores for an average group of students may have been expected to be about .4; however, the total reading grade equivalent gains for these students, who presumably had never made average gains in their previous years of education, averaged .6 across the three tests.

Figure 4
Comparisons of Pretest/Posttest Scores on the SAT 9
Using Grade 7 National Norm Comparisons



Grade 8 Results: Stanford Achievement Test, Primary 3

For Grade 8 students, the average standard scores increased from pretest to posttest for Reading Comprehension and Total Reading. However, the increases were not statistically significant for Reading Vocabulary as can be seen in Table 3. The increases in Reading Comprehension and Total Reading were significant beyond the .01 level. This means that such changes in scores would be expected to happen by chance fewer than once out of a hundred times.

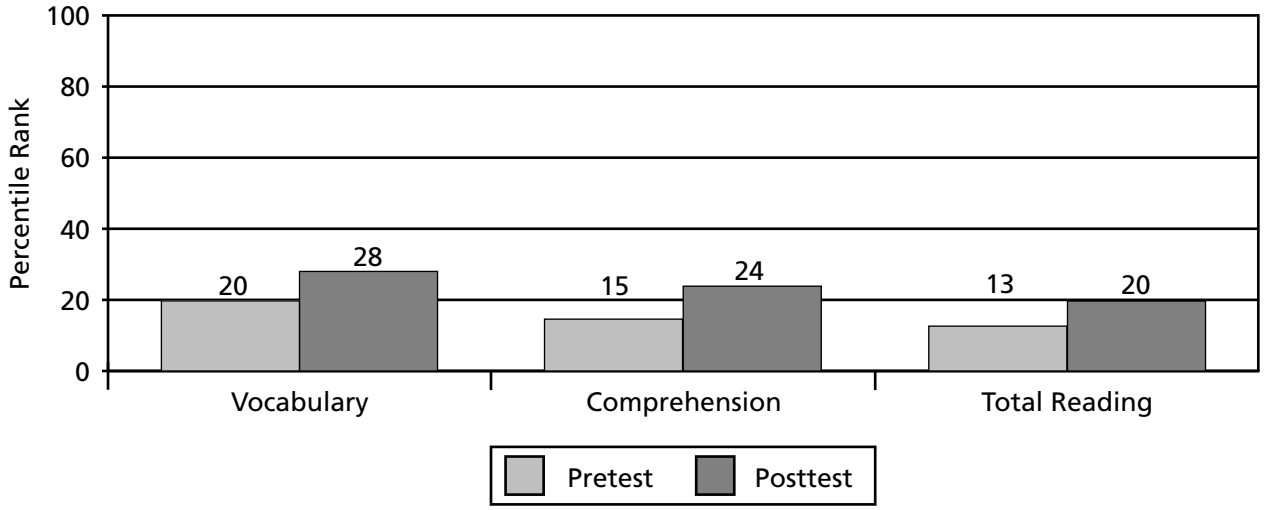
The lack of a statistically significant increase for Reading Vocabulary may have been because the Grade 8 students scored at relatively high levels on both the pretests and the posttests. The average raw scores for the students on the 20-item Vocabulary were 17.00 on the pretest and 17.24 on the posttest. As with the 6th grade students, the Primary 3 level Vocabulary test of the SAT 9 may have been at too low a level for these students. However, the comprehension scores indicate that the test was at an appropriate level. Since the *Power Up!* program focuses primarily on reading comprehension, the test seems to have been an appropriate choice overall. Although pre/posttest scores moved in a positive direction, given the small effect size for all three scores, the magnitude of the result is not very strong.

Table 3
Paired Comparison *t*-test Gains from Pretest to Posttest
for Grade 8 Students on the SAT 9, Abbreviated
N=53

Test Score	Pretest Standard Scores		Posttest Standard Scores		Paired Comparison *t*-test		Effect Size	
	Mean	*Standard Deviation*	*Mean*	*Standard Deviation*	*t-test*	*Significance*	*Cohen's d*	*Effect Size r*
Vocabulary	655.62	44.34	658.09	39.81	.42	<.49	.058	.029
Comprehension	637.09	51.50	651.67	46.06	2.675	<.01	.298	.147
Total	641.54	42.06	652.34	38.48	3.122	<.01	.267	.132

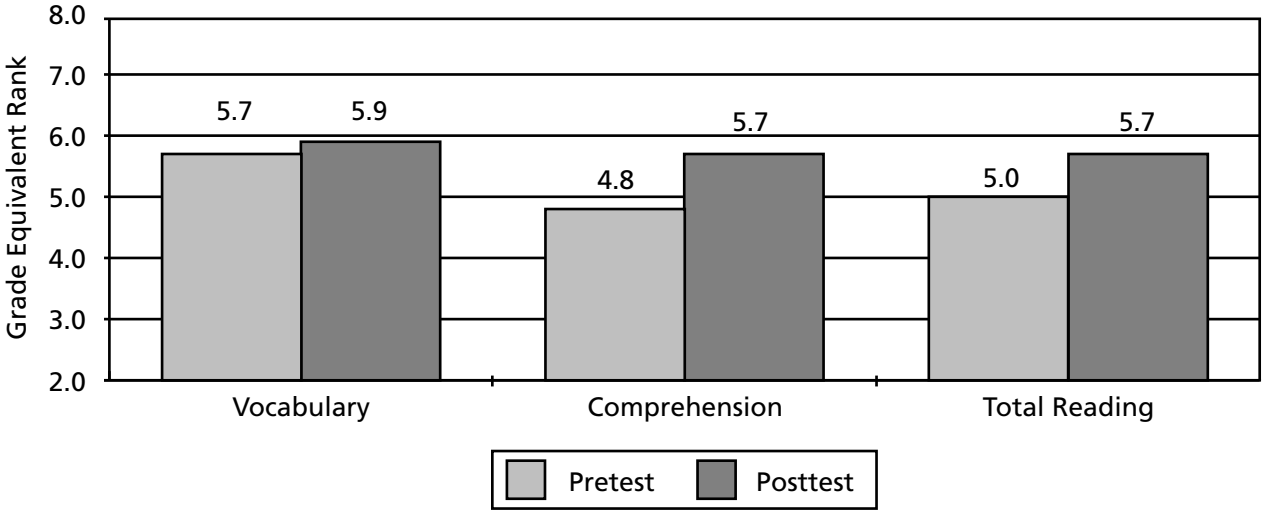
The percentile comparisons of the pretest to posttest gains are shown in Figure 5. These percentile comparisons are based on Grade 8 norms, as these students were enrolled in eighth grade at the time of the study. The SAT 9 provides norm tables that describe how Grade 8 students performed nationally on the Primary 3 level of the SAT 9 Abbreviated. The standard score analysis for Grade 8 students did reveal statistically significant gains, as shown in Table 3. These gains can also be seen in Figure 5 in terms of percentile increase; students averaged an 8 percentile point increase across the three scores.

Figure 5
Comparisons of Pretest/Posttest Percentile Scores on the SAT 9, Abbreviated Test Using Grade 8 National Norm Comparisons



The SAT 9 also provides grade equivalent scores to interpret the test on a norm-referenced basis. The gains in grade equivalent scores for the Grade 8 students are shown in Figure 6. Figure 6 provides further evidence about the low reading levels of the Grade 8 students at the beginning of the study. It appears that these eighth grade students were reading about two years below grade level at the beginning of the study. The comparison of these grade equivalent gains to the national averages is, however, significant. Since the study took place over a 4-month period, the gain scores for an *average* group of students may have been expected to be about .4; however, the total reading grade equivalent gains for these students, who presumably had never made average gains in their previous years of education, were .9 for comprehension (almost a full grade level) and .7 for Total Reading. Reading Vocabulary only showed a .2 gain. However, the observed ceiling effect likely prevented much gain with the vocabulary test scores.

Figure 6
Comparisons of Pretest/Posttest Grade Equivalent Scores on the SAT 9
Using Grade 8 National Norm Comparisons



Summary

Table 4 below summarizes the results across grade levels. The first three columns provide the pretest and posttest grade eqivalent scores for all three subtests for Grade 6, 7, 8 students. The fourth column shows what the average grade equivalent gain was for a similar period for students in the national norm population. The final column indicates whether the gains were statistically significant. The statistical significance was based on the *t*-tests computed using the standard scores for each of the tests.

Table 4
Summary of Grade Equivalent Gain/Losses
on the Stanford Achievement Test: Primary 3
for Grade 6, 7, and 8 Students

	Grade Equivalent Gains				Standard Score *t*-test gains
Grade/Test	Pretest Grade Equivalent	Posttest Grade Equivalent	Change from Pre to Post (+/-)	Expected Gain Based on National Norms	Significance
Grade 6					
Vocabulary	5.8	6.5	+ .7	+ .4	No
Comprehension	5.2	6.1	+ .9	+ .4	Yes
Total	5.6	6.7	+1.1	+ .4	Yes
Grade 7					
Vocabulary	4.3	4.7	+ .4	+ .4	Yes
Comprehension	3.7	4.5	+ .8	+ .4	Yes
Total	3.9	4.5	+ .6	+ .4	Yes
Grade 8					
Vocabulary	5.7	5.9	+ .2	+ .4	No
Comprehension	4.8	5.7	+ .9	+ .4	Yes
Total	5.0	5.7	+ .7	+ .4	Yes

Conclusions

1. The students participating in this efficacy study of *Power Up!: Building Reading Strength* experienced achievement gains on the SAT 9 that increased as much or more than students in the national norm population at all three grade levels for Reading Vocabulary, Reading Comprehension, and Total Reading.

2. Statistical analyses of the standard scores for each of the tests revealed that achievement gains were statistically significant for Reading Comprehension and Total Reading at all three grade levels. Gain scores for Reading Vocabulary were statistically significant at grade 7, but not at grades 6 and 8. A ceiling effect (high scores on pretests) may have prevented the gain scores from achieving statistical significance at these grades.

3. Calculated effect sizes between pre/posttest means were in a positive direction on all tests at all grade levels. The magnitude of the effect was stronger for Reading Comprehension and Total Reading at the 6th grade than it was for the other scores and grade levels. For all other scores and grade levels, the strength of the effect was considered to be small; however, given the short duration of the study such effect sizes are not unexpected.

4. The study lasted only about four months. The program was new to the teachers and they received only one in-service session with a consultant. Given the results of this study, it can be assumed that greater gains would likely result when students use the *Power Up!* program for an entire school year and/or when teachers are more familiar with the program.

Notes

Rigby • Steck-Vaughn

P.O. Box 690789
Orlando, FL 32819-0789
1-800-531-5019 • www.steck-vaughn.com

5M/HAGG/AAP/10-03
9993988650 INDIV
9993988669 25-PACK