

**A STUDY OF FLOCABULARY'S *THE WORD UP PROJECT*
PROGRAM'S INFLUENCE ON STATE
READING/LANGUAGE ARTS TEST SCORES: A
TREATMENT AND CONTROL GROUP DESIGN**

Report 372
October 2009

Advisory Board:

Michael Beck, President
Beck Evaluation & Testing Associates, Inc.

Jennifer M. Conner, Assistant Professor
Indiana University

Keith Cruse, Former Managing Director
Texas Assessment Program



Table of Contents

Table of Contents	1
Purpose of the Study	2
Research Question	3
Design and Procedures of the Study	3
The Instructional Program	5
The State Assessments Used for Comparative Analyses.....	6
The Study Sample	8
Schools Included in the Study.....	8
Data Analyses: Academic Achievement Equal	9
School 1: Pennsylvania.....	9
School 2: Alabama.....	10
School 3: New York.....	12
School 4: Massachusetts	13
School 5: Texas.....	14
Data Analyses: Academic Achievement Unequal	16
School 6: Texas.....	16
School 7: California	17
School 8: Texas.....	19
Summary of Results	21

A Study of Flocabulary’s *THE WORD UP PROJECT* Program’s Influence on State Reading/Language Arts Test Scores: A Treatment and Control Group Design

This report describes the study of the instructional effectiveness of *THE WORD UP PROJECT*, a vocabulary program designed for students in grade 3 to 8.

Purpose of the Study

Flocabulary, an educational publisher, contracted with the Educational Research Institute of America (ERIA) to conduct an instructional validation study of its grades 3 to 8 language arts program, *THE WORD UP PROJECT*, in eight school sites across the country.

Flocabulary sought to determine the extent to which *THE WORD UP PROJECT*, a program designed to support students’ vocabulary development, reading comprehension, and critical thinking skills, improves students’ language arts skills as measured by state language arts tests.

The study included eight different schools in six different states – Alabama, California, Massachusetts, New York, Pennsylvania and Texas. All of the students in the study were enrolled in grades 6, 7, or 8. Students began using *THE WORD UP PROJECT* in September of 2008 and continued until May of 2009. A pretest and posttest designed to measure vocabulary development was administered to students prior to and upon their completion of the program. In addition, student scores on each state’s end-of-year language arts test were provided by each school for students using *THE WORD UP PROJECT*, as well as for a control group in each school.

The results of the analysis of *THE WORD UP PROJECT* students’ pretest and posttest scores on the vocabulary development test are included in a companion report. The study reported here provides an analysis of the data comparing state language arts test scores for the students in each school using *THE WORD UP PROJECT* with a control group at each school. Each control group consisted of students enrolled in the same grade level as the students at that school who used *THE WORD UP PROJECT*. However, the students in the control groups did not use *THE WORD UP PROJECT*.

Research Question

The following research question guided the design of the data analyses:

Is *THE WORD UP PROJECT* instructionally effective and do students who receive instruction based on the program's teacher's edition and student text demonstrate significantly stronger reading and language arts skills than a control group as measured by state language arts assessments?

Design and Procedures of the Study

Eight schools in six different states agreed to participate in the study. One or more teachers in each of the eight schools agreed to use *THE WORD UP PROJECT* during the 2008-2009 academic year. The goal was to determine if the reading and language arts skills of students who received instruction using *THE WORD UP PROJECT* would be stronger than an academically similar control group by the end of the school year.

The test scores used to determine reading and language arts achievement came from the state achievement tests in each of the five states. Since those tests are not administered until near the end of the school year, test results were not available to ERIA until the summer and fall of 2009.

A control group comprised of students who did not use *THE WORD UP PROJECT* but who were at the same grade levels as students who did was provided by each of the schools. These control group students were to be selected so that they were approximately equal in academic achievement to *THE WORD UP PROJECT* students. However, questionnaire surveys completed by teachers or administrators in each of the schools indicated that in only three of the schools did the academic achievement of the students in *THE WORD UP PROJECT* group equal the academic achievement of the control group students. In two other schools the academic achievement of the control group students was somewhat higher than that of the students in *THE WORD UP PROJECT* group, but not to a large extent. However, in three of the schools, the academic achievement of the control group significantly exceeded that of the students in *THE WORD UP PROJECT* group.

Differences in the academic achievement between the two groups were determined from the survey responses provided by each school. For example, one school reported large academic achievement discrepancies between *THE WORD UP PROJECT* classes and the control classes. Their survey response stated that *THE WORD UP PROJECT* classes included a number of special education students, no gifted and talented students, and all students in *THE WORD UP PROJECT* classes were reading two years or more below grade level. On the other hand, the school survey results indicated that the students in the control group classes included no special education students, a number of gifted and talented students, and fewer students reading two or more years below grade level. The control group in this school did not comprise a fair comparison.

In this school and for two others with similar large discrepancies, a control sample was selected from the total submitted control group that was equal in size to the number of students in *THE WORD UP PROJECT* classes. The selected control groups were

comprised of those students whose state test scores were the lowest in the original sample. *While this adjustment does not provide a true control group comparison, it is possible that it is closer to a satisfactory comparison than the original control group.*

Thus, two different analyses were conducted as follows:

1. When the students in *THE WORD UP PROJECT* group and the control group classes were approximately equal (three schools) or the control groups were not significantly higher in terms of academic achievement (two schools) no adjustments were made with the data.
2. When the students in the control classes were significantly higher than the students in *THE WORD UP PROJECT* group (three schools), a sample of students was selected whose state scores were at the lowest levels of the control group. This selection provided a control group that was perhaps closer in achievement to *THE WORD UP PROJECT* students.

All of the state test scores for both groups were submitted to ERIA by each of the eight schools. The state test scores were then compared using a one-way analysis of variance. The $<.05$ significance level was used to reject or accept whether the difference between the mean scores of each group was statistically significant.



The Instructional Program

The description of *THE WORD UP PROJECT* provided by the publisher's website describes the program as follows:

THE WORD UP PROJECT is a multisensory instructional program designed to build vocabulary proficiency and reading skills. It combines rigorous research with a standards-based curriculum, and features high-interest music and exercises. The program teaches vocabulary, reading comprehension, and critical thinking in an engaging format. The vocabulary words in each level have been selected according to how frequently they appear on state assessments.

The student books and CDs provide multiple and varied exposures to each word, features exercises that mimic those on many high-stakes tests, and empowers students to use the words in their own speech and writing. The Teacher Resource Book includes everything an educator needs to implement the program in her classroom. The program is now available in six reading levels: 3rd grade (red), 4th grade (orange), 5th grade (indigo), 6th grade (green), 7th grade (yellow) and 8th grade (blue).

The State Assessments Used for Comparative Analyses

The information presented below provides a brief overview of the assessments used in each state and which were used for this study. All of the information was taken from the Department of Education websites for each of the states. Some editing was done to provide consistency across descriptions. For more information about the assessments, the Department of Education websites provide extensive descriptions of their assessments.

Alabama

The Alabama Reading and Mathematics Tests (ARMT) are administered to all students in grades 3 through 8. The ARMT test is based on the Alabama state standards, which define what students should be learning each year. The test used as the ARMT reading assessment is the Stanford Achievement Test Tenth Edition (SAT10). The SAT10 is used to compare individual and group performance with performance of the norming group, to report relative strengths and weaknesses of individuals and groups, and to provide data to study changes in performance over time

California

The California State Tests (CSTs) are a major component of the Standardized Teaching and Reporting (STAR) program. The CSTs are developed by California educators and test developers specifically for California. They measure students' progress toward achieving California's state-adopted academic content standards, which describe what students should know and be able to do in each grade and subject tested.

Massachusetts

Massachusetts Comprehensive Assessment System—English Language Arts (MCAS-ELA) assesses reading and writing skills across all grade levels. The MCAS tests are based on the learning standards in the Massachusetts Curriculum Frameworks. There are twelve key language, reading, and literature standards that are the basis for the MCAS-ELA assessments. These include Vocabulary and Concept Development, and Formal and Informal English. MCAS tests include multiple-choice questions, short-answer questions, open-response questions, and writing prompts.

New York

The New York State English Language Arts (NYSELA) tests are administered grades 6, 7, and 8. The English Language Arts Tests consist of a section containing multiple-choice and short-response questions based on reading selections and a section containing multiple-choice and short-response questions based on a listening selection. The second section also contains an editing task.

Pennsylvania

The Pennsylvania System of School Accountability (PSSA) applies to all public schools and districts. The PSSA assessments are based on the State's content and achievement standards. The PSSA includes a reading assessment for grades 3 to 8 and 11. Math and reading are the only two tests that count towards a school's Adequate Yearly Progress (AYP) status.

Texas

The Texas Assessment of Knowledge and Skills (TAKS) reading assessment is administered each year to students in grades 3 to 11. The TAKS reading assessments evaluate a subset of the Texas Essential Knowledge and Skills (TEKS), the state-mandated curriculum. This curriculum is specifically designed to help students make progress in reading by emphasizing the knowledge and skills most critical for student learning. Because the TAKS reading tests are closely aligned with the TEKS, students who effectively learn the TEKS will become proficient readers who are able to perform successfully on the test without unnecessary emphasis on test preparation.

The Study Sample

Eight schools in six different states were included in the study. A total of 572 students made up the total group in *THE WORD UP PROJECT* classes, and 683 students were included in the control group classes.

Schools Included in the Study

Table 1 shows the demographic data for each of the eight schools. On average, 53% of students across the eight schools were enrolled in free/reduced lunch programs, ranging from 20% to 87%. An average of 52% of the students across the study schools were considered to be minority students, ranging from 4% to 100%. The average enrollment at the schools was 890 students, ranging from 684 to 1,499 students.

Table 1
Demographic Characteristics of the Schools Included in the Study

<i>School</i>	<i>State</i>	<i>Locale</i>	<i>Grades</i>	<i>Enrolment</i>	<i>% Free Lunch</i>	<i>% Minority</i>
1	PA	Small Town	6 to 8	684	20%	4%
2	AL	Small Town	5 to 8	686	51%	27%
3	NY	Large Central City	K to 8	1499	68%	74%
4	MA	Urban Fringe Of Mid-Sized City	6 to 8	764	40%	17%
5	TX	Large Central City	6 to 8	638	87%	100%
6	TX	Urban Fringe Of Large City	6 to 8	980	48%	79%
7	CA	Urban Fringe Of Large City	6 to 8	1147	63%	74%
8	TX	Urban Fringe Of Large City	6 to 8	718	43%	40%
<i>Average</i>				890	53%	52%

Data Analyses: Academic Achievement Equal

Five schools (schools 1, 2, 3, 4 and 5) reported that the students in *THE WORD UP PROJECT* classes and the control classes were approximately equal in academic achievement.

School 1: Pennsylvania

A comparison of the performance of *THE WORD UP PROJECT* students (N=26) to a control group of students (N=26) was conducted using the end-of-year Pennsylvania System of School Assessment: Reading (PSSA) test scores. Both *THE WORD UP PROJECT* students and the control group students were chosen from a range of academic achievement level groups. The goals of the selection process were to have two groups that were approximately equal in achievement level representation, and that were representative of the school as a whole. School questionnaire results indicate that the teachers felt that both goals were achieved.

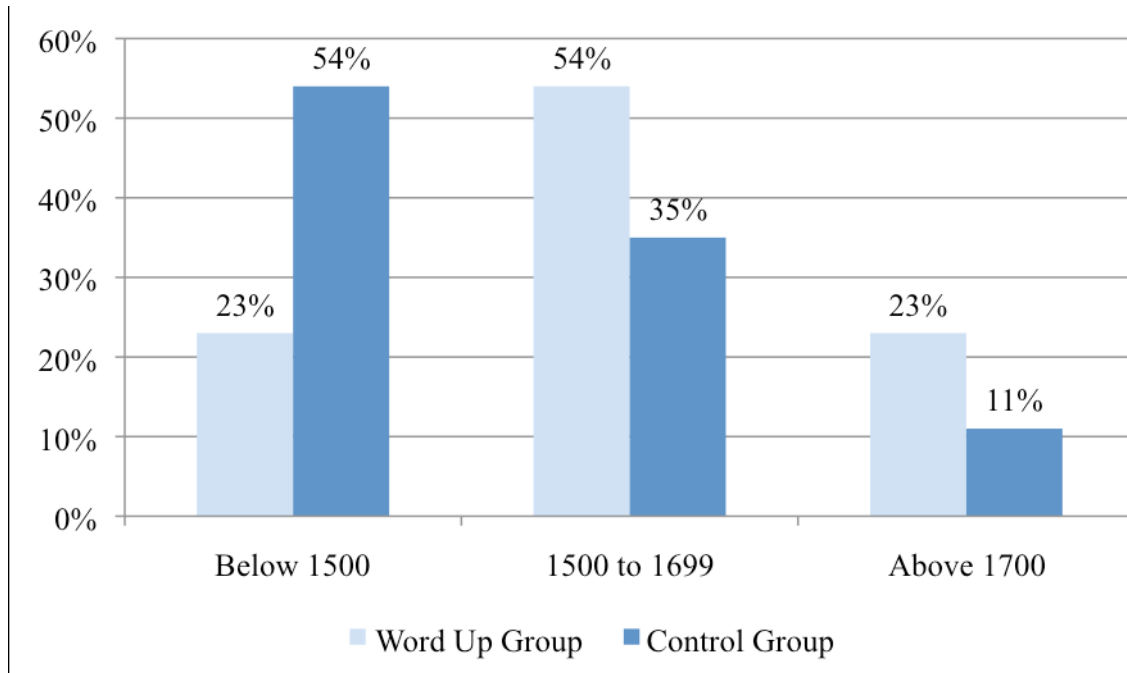
Table 2 shows the analysis of variance comparison for the groups. The results indicate that the average score of *THE WORD UP PROJECT* students were statistically significantly higher (<.01) in comparison to the control group students on the PSSA Reading. This level of significance indicates that such a difference would occur by chance fewer than once out of 100 repetitions.

Table 2
Analysis of Variance Comparing *THE WORD UP PROJECT* Group to a Control Group on the Pennsylvania System of School Assessments: Reading (PSSA) Test at School 2

	N	Mean Score	Standard Deviation	F Test	Significance
Word Up Group	26	1615	223	7.091	<.004
Control Group	26	1452	219		

Figure 1 shows that fewer students in *THE WORD UP PROJECT* group (23%) scored at the lower levels than did the control group students (54%). In addition, more *THE WORD UP PROJECT* students (23%) scored at the higher level than did the control group students (11%).

Figure 1
Percent of *THE WORD UP PROJECT* and Control Group Students
Scoring at Various Score Levels on the
Pennsylvania System of School Assessments: Reading (PSSA) Test at School 3



School 2: Alabama

The response to the survey question about the academic achievement of the two groups follows:

The students in both groups were a mix of below grade level, on grade level, and above grade level. My 7th grade Enrichment class has all students except those who choose to be in band/choir or yearbook. There is a small group pulled out for a reading enrichment class – about 10 students; however, there are still students who are below grade level in reading/writing and students with IEPs/504s in my inclusion classroom. Both groups are a good representation of different academic and socioeconomic groups in our community. They really aren't different.

A comparison of the performance of *THE WORD UP PROJECT* students (N=82) to a control group of students (N=98) was conducted using the end-of-year Alabama Reading and Math Test (ARMT). The reading test used for the reading portion of the ARMT is the

Stanford Achievement Test. Results are provided as standard scores. In addition, the results provide a score for each student ranging from 1 to 4. These scores represent four levels of achievement:

- 1 Does Not Meet Academic Standards
- 2 Partially Meets Academic Standards
- 3 Meets Academic Standards
- 4 Exceeds Academic Standards

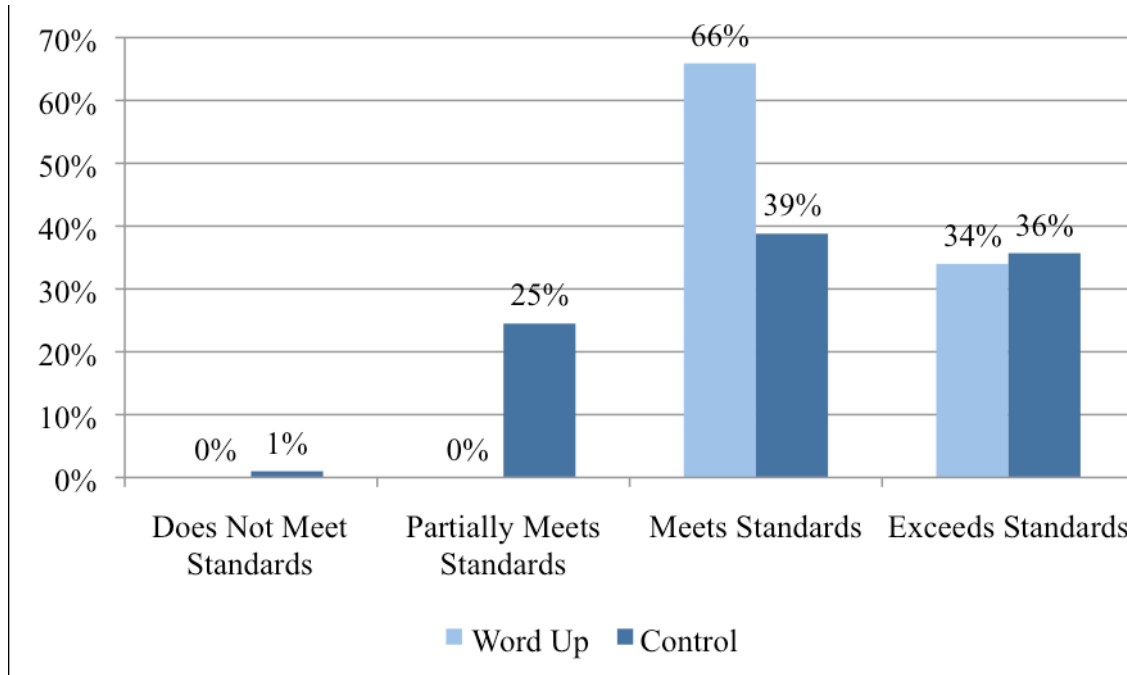
Table 3 shows the analysis of variance comparison for the reading portion of the ARMT for the two groups. The results indicate that the average score of *THE WORD UP PROJECT* students was significantly higher (<.002) than that of the control group students. This level of significance indicates that such a difference would occur by chance less than twice out of 1,000 repetitions.

Table 3
Analysis of Variance Results Comparing *THE WORD UP PROJECT* Group to a Control Group on the Reading Portion of the Alabama Reading and Math Test (ARMT) at School 1

	N	Mean Score	Standard Deviation	F Test	Significance
Word Up Group	82	681.2	26.8	9.720	<.002
Control Group	98	669.7	34.8		

A comparison of the percentage of *THE WORD UP PROJECT* and control group students scoring at each of the four levels on the reading portion of the ARMT administered at the end of the school year are shown in Figure 2. The results show that no *THE WORD UP PROJECT* student was at the *Does Not Meet Standards* or *Partially Meets Standards* at the end of the academic year. One hundred percent of *THE WORD UP PROJECT* students scored at the *Meets* or *Exceeds* standards levels.

Figure 2
Percent of *THE WORD UP PROJECT* Students and Control Group Students
Scoring at Various Score Levels on the
Reading Portion of the Alabama Reading and Math Test (ARMT) at School 1



School 3: New York

A comparison of the performance of *THE WORD UP PROJECT* students (N=89) to a control group of students (N=119) was conducted using the end-of-year New York State English Language Arts (NYSELA) test scores. Both *THE WORD UP PROJECT* students and the control group students were chosen from high, average, and low academic achievement level groups. The goals of the selection process were to have two groups that were approximately equal in achievement level representation, and that were representative of the school as a whole. School questionnaire results indicate that the teachers felt that both goals were achieved.

Table 4 shows the analysis of variance comparison for the groups. The results indicate that the average score of *THE WORD UP PROJECT* students increased statistically significantly (<.004) in comparison to the control group students on the NYSELA tests. This level of significance indicates that such a difference would occur by chance fewer than four times out of 1,000 repetitions.

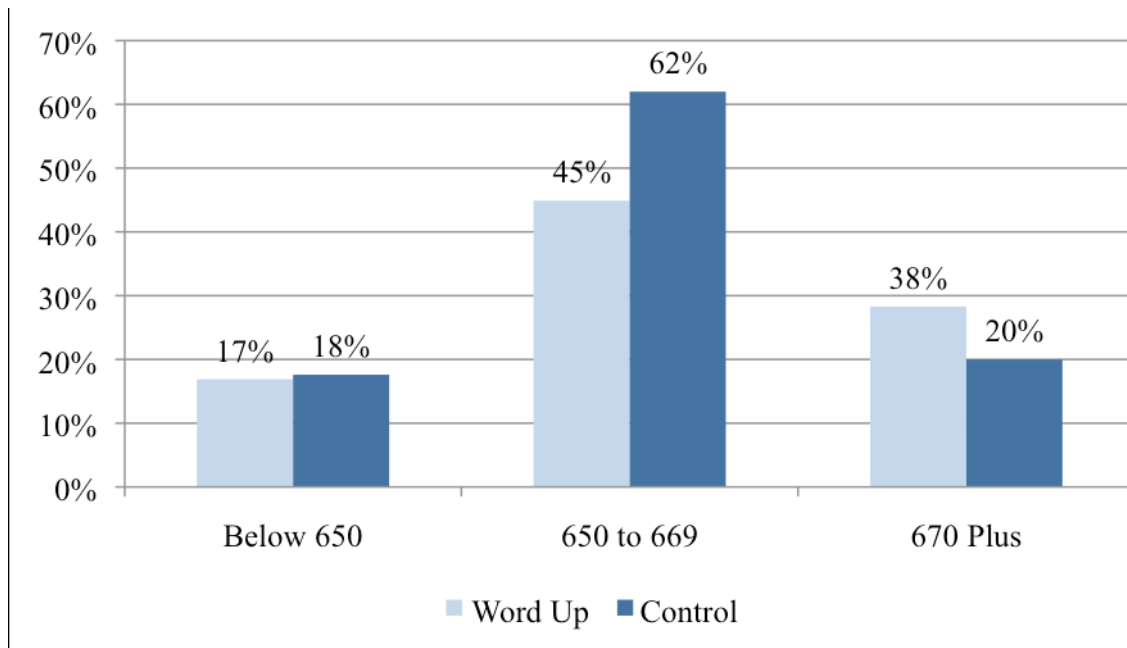
Table 4
Analysis of Variance Comparing *THE WORD UP PROJECT* Group to a Control
Group on the New York State English Language Arts (NYSELA) Test at School 3

	N	Mean Score	Standard Deviation	F Test	Significance

Word Up Group	89	669	28.4	8.350	<.004
Control Group	119	660	15.6		

Figure 3 shows that fewer students in *THE WORD UP PROJECT* group (62%) scored at the lower levels than did the control group students (45%). In addition, more *THE WORD UP PROJECT* students (80%) scored at the higher level than did the control group students (20%).

Figure 3
Percent of *THE WORD UP PROJECT* and Control Group Students Scoring at Various Score Levels on the New York State English Language Arts (NYSELA) Test at School 3



School 4: Massachusetts

A comparison of the performance of *THE WORD UP PROJECT* students (N=85) to a control group of students (N=141) was conducted using the end-of-year Massachusetts Comprehensive Assessment System—English Language Arts (MCAS-ELA) standard scores.

Questionnaire results from the school indicated that the students in *THE WORD UP PROJECT* groups were in regular language arts classes. However, the students at this grade are assigned to language arts classes based on whether they were enrolled in band or chorus classes. *THE WORD UP PROJECT* group was the group enrolled in chorus and the control group was the group that was enrolled in band. The teachers reported that the students in the band group generally perform somewhat better than the chorus group students. Thus, one might expect *THE WORD UP PROJECT* students to perform somewhat lower on the MCAS-ELA assessments. However, that turned out not to be the case. In fact, both groups performed almost exactly the same on the state test.

Table 5 shows the analysis of variance comparison on the MCAS-ELA for the two groups. The results indicate that the average scores of *THE WORD UP PROJECT* students and the control group students were not statistically different.

Table 5
Analysis of Variance Results Comparing *THE WORD UP PROJECT* Group to a Control Group on the Massachusetts Comprehensive Assessment System—English Language Arts (MCAS-ELA) Test at School 4

	N	Mean Standard Score	Standard Deviation	F Test	Statistical Significance
Word Up Group	85	240.8	11.2	.006	None
Control Group	141	240.9	14.5		

School 5: Texas

A comparison of the performance of *THE WORD UP PROJECT* students (N=84) to a control group of students (N=93) was conducted using the end-of-year Texas Assessment of Knowledge and Skills (TAKS) reading assessment standard scores.

Questionnaire results from the school indicated that the students in *THE WORD UP PROJECT* groups were in regular classes and no pre-advance placement classes were included in those classes. However, the questionnaire results indicated that the control group included both regular class students and pre-advance placement students. It is not known how many advance placement students were included in the control group.

Table 6 shows the analysis of variance comparison on the TAKS reading assessment between the two groups. The results indicate that the average scores of *THE WORD UP PROJECT* students and the control group students were not statistically different. The lack of any statistical significance indicates that any difference would be considered a chance difference.

Table 6
Analysis of Variance Results Comparing *THE WORD UP PROJECT* Group to a Control Group on the Texas Assessment of Knowledge and Skills (TAKS) Reading Assessment at School 5

	N	Mean Standard Score	Standard Deviation	F Test	Statistical Significance
Word Up Group	84	2205.7	138	3.411	None
Control Group	93	2245	145		

Data Analyses: Academic Achievement Unequal

There were three schools (schools 6, 7, and 8) that reported that the students in *THE WORD UP PROJECT* group were unequal in academic achievement to the control group students. The academic differences in each of these three schools were very much in favor of the control group students. At each of these schools, a sample was selected from the total control group that was equal in size to the number of students in *THE WORD UP PROJECT* classes. The adjusted control groups were comprised of those students whose state test scores were the lowest in the original sample.

School 6: Texas

A comparison of the performance of *THE WORD UP PROJECT* students (N=39) to a control group of students (N=39) was conducted using the end-of-year Texas Assessment of Knowledge and Skills (TAKS) reading assessment standard scores.

Questionnaire results from the school indicated that the students in *THE WORD UP PROJECT* were all enrolled in classes established for struggling readers. None of the control group classes were considered classes for struggling readers.

Since *THE WORD UP PROJECT* students almost certainly represented students who in the past had scored at the lowest levels on the TAKS test, it was decided that a fairer comparison would be to compare the 39 *THE WORD UP PROJECT* students to the 39 students in the control group with the lowest state test scores.

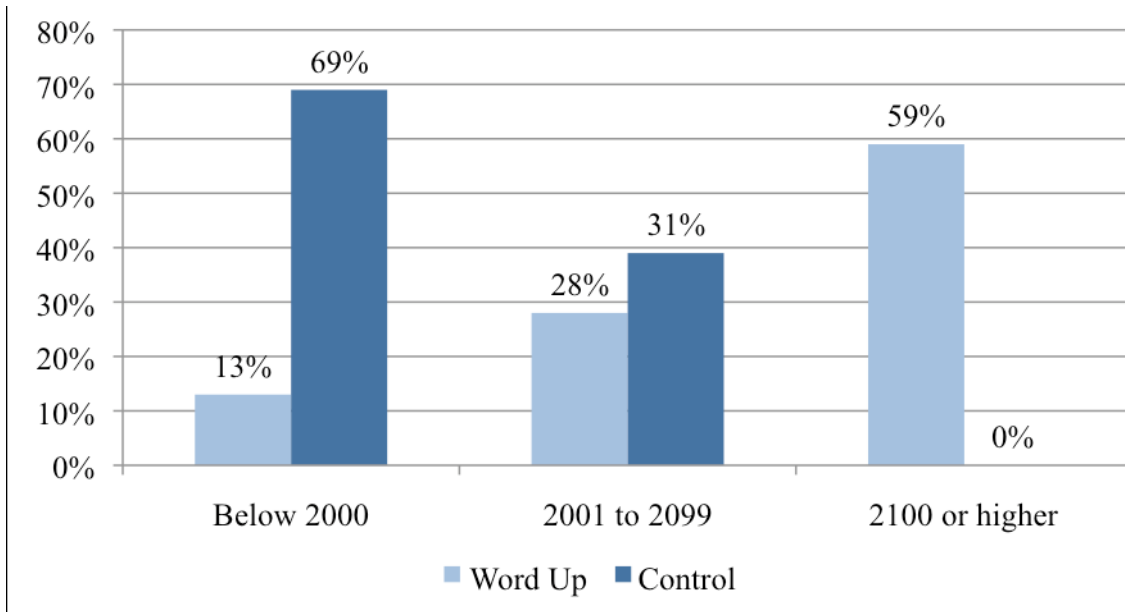
Table 7 shows the analysis of variance comparison for the TAKS reading test for the two groups. The results indicate that the average scores of *THE WORD UP PROJECT* students were significantly higher than the control group students (<.0001), a statistical level that indicates a difference that would occur by chance less than once out of 10,000 repetitions.

Table 7
Analysis of Variance Results Comparing *THE WORD UP PROJECT* Group to a Control Group on the Texas Assessment of Knowledge and Skills (TAKS) Reading Assessment at School 6

	N	Mean Standard Score	Standard Deviation	F Test	Statistical Significance
Word Up Group	39	2109.7	116.6	50.326	<.0001
Control Group	39	1943.1	88.9		

Figure 4 below compares the percentage of control and *THE WORD UP PROJECT* students scoring below 2000 on the test, from 2001 to 2099, and 2100 or higher. Fifty-nine percent of *THE WORD UP PROJECT* students scored at the highest level on the TAKS while none of the control group students scored at this level. Sixty-nine percent of the control group students scored at the lowest level while only 13% of *THE WORD UP PROJECT* students scored at the lowest level.

Figure 4
Percent of *THE WORD UP PROJECT* and Control Group Students
Scoring at Various Score Levels on the Texas Assessment of Knowledge and Skills
(TAKS) Reading Assessment at School 6



School 7: California

A control group was compared to the performance of *THE WORD UP PROJECT* on the California STAR assessment. The test score used for comparison was the STAR vocabulary score which was available as a percent correct score.

The students’ state test scores were provided by the school. STAR vocabulary scores were provided for 64 of *THE WORD UP PROJECT* students. The control group included a total of 227 students.

The two groups were very different in terms of academic achievement. Surveys sent to the schools were used to gather data about the academic levels of *THE WORD UP PROJECT* students and the control group students. The surveys revealed the following differences between the two groups:

THE WORD UP PROJECT group was academically less able compared to the control group. Seventy-five percent of the students in THE WORD UP PROJECT group were classified as English Language Learners while only 40% of the control group students were classified as English Language Learners.

In addition, THE WORD UP PROJECT group included a large number of students reading 1.5 to 2.0 years or more below grade level. One of the classes included 19 out of a total of 28 students who were classified as special education. None of the students in the control group were classified as special education.

The control group included some students reading two years or more below grade level and none of the students were classified as special education. The control group also included all of those students classified as gifted and talented at the grade level included in the study.

Since *THE WORD UP PROJECT* students almost certainly represented students who in the past had scored at the lowest levels on the STAR test, it was decided that a fairer comparison would be to compare the 64 *THE WORD UP PROJECT* students to the 64 students in the control group with the lowest state test scores.

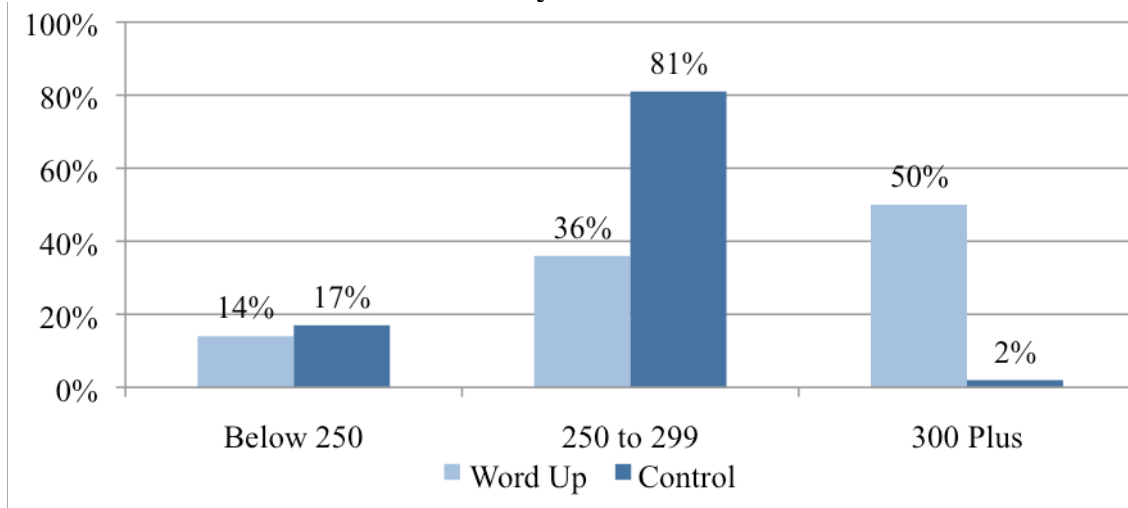
Table 8 shows the analysis of variance comparison of the STAR vocabulary scores for the two groups. The results indicate that the average scores of *THE WORD UP PROJECT* students were significantly higher than the control group students (<.0001), a statistical level that indicates a difference that would occur by chance less than once out of 10,000 repetitions.

Table 8
Analysis of Variance Results Comparing *THE WORD UP PROJECT* Group to a Control Group on the California STAR Vocabulary Test at School 7

	N	Mean Standard Score	Standard Deviation	F Test	Statistical Significance
Word Up Group	64	45.8	20.6	43.616	<.0001
Control Group	64	26.6	10.5		

Figure 5 below compares the percentage of control and *THE WORD UP PROJECT* group students scoring below 250, from 250 to 299, and 300 or higher on the STAR vocabulary test. Ninety-eight percent of the control group students scored at the lowest two levels while only 50% of *THE WORD UP PROJECT* students scored at this level. On the other hand, 50% of *THE WORD UP PROJECT* students scored at the highest level while only 2% of the control group students scored at this level.

Figure 5
Percent of *THE WORD UP PROJECT* and Control Group Students
Scoring at Various Score Levels on the California
STAR Vocabulary Assessment at School7



School 8: Texas

A comparison of the performance of *THE WORD UP PROJECT* students (N=103) to a control group of students (N=121) was conducted using the end-of-year Texas Assessment of Knowledge and Skills (TAKS) reading assessment standard scores.

Questionnaire results from the school indicated that the students in *THE WORD UP PROJECT* group were in general academically weaker than the students in the control group. The following statement was submitted by the school:

THE WORD UP PROJECT was completed by on-level students who are not in resource classes. The segment of our school that did not participate falls into three categories: resource, Pre-AP and Gifted & Talented.

Since *THE WORD UP PROJECT* students almost certainly represented students who in the past had scored at the lowest levels on the TAKS test, it was decided that a fairer comparison would be to compare the 103 *THE WORD UP PROJECT* students to the 103 students in the control group with the lowest state test scores.

Table 9 provides the results of an analysis of variance comparison between the two groups and shows that the average scores of *THE WORD UP PROJECT* students and the control group students were not statistically different. The control students scored the same on the TAKS test as *THE WORD UP PROJECT* students.

Table 9
Analysis of Variance Results Comparing *THE WORD UP PROJECT* Group to a Control Group on the Texas Assessment of Knowledge and Skills (TAKS) Reading Assessment at School 8

	N	Mean Standard Score	Standard Deviation	F Test	Statistical Significance
Word Up Group	103	2227	137	.0000	Non-Significant
Control Group	103	2227	168		

Summary of Results

Following is a summary of the results of the analyses comparing the state test scores of *THE WORD UP PROJECT* students compared to control group students. A total of eight schools in six states were included in the study.

The data analysis divided the schools into two sets of schools. This division was based on whether the previous academic achievement of *THE WORD UP PROJECT* students and the control group students was relatively equal or not. It was planned that the schools would provide state test scores for control group students whose academic achievement was approximately equal to *THE WORD UP PROJECT* students. This request was followed for five schools. However, the remaining three schools provided state test scores for control group students whose previous academic achievement was significantly higher than *THE WORD UP PROJECT* students.

Thus, the first set of five schools included those schools in which the students' previous academic achievement was about equal. The second set of schools were those in which the schools provided control group students whose previous academic achievement was much higher than *THE WORD UP PROJECT* students, thus providing a control group that would not result in a valid comparison. To attempt to work around this problem, the control group students in the second group of schools were selected based on their performance on the state test scores. Those students whose scores were the lowest on the state tests were selected for the statistical analyses.

It is important to note that neither the original control group provided by the three schools nor the group selected from that sample by the researchers provide a true comparison group.

The summary of the results in Table 10 show that for the five schools in which the schools provided comparable control group students, *THE WORD UP PROJECT* students scored significantly higher than the control group students in 3 of the schools. In the other 2 schools, there was a slight advantage in previous academic achievement for the control group students. In those 2 schools there was not statistically a significant difference between *THE WORD UP PROJECT* students and the control group students.

For the 3 schools in which there was a significant academic achievement advantage for the control group students, a group was selected by researchers which may have provided a more equitable control group. The comparison between *THE WORD UP PROJECT* students in 2 of these schools provided a statistically significant comparison in favor of *THE WORD UP PROJECT* students. In the third school, there was no significant difference.

Table 10
Summary of Results Comparing Control Group Students to *THE WORD UP PROJECT* Group Students on State English Language Arts Tests

School	Differences in Academic Achievement Between Groups	Higher Group on State Test Results	Statistical Significance
<i>Groups Not Adjusted</i>			
Pennsylvania	None	Word Up	<.01
Alabama	None	Word Up	<.002
New York	None	Word Up	<.004
Massachusetts	Control group slightly higher	Neither	Non-Significant
Texas	Control group slightly higher	Neither	Non-Significant
<i>Groups Adjusted for Academic Differences</i>			
Texas	Control group much higher	Word Up	<.0001
California	Control group much higher	Word Up	<.0001
Texas	Control group much higher	Neither	Non-Significant

The results from eight schools provide significant evidence that students who used THE WORD UP PROJECT receive higher scores on state reading and language arts scores than students who did not use the program. The lack of adequate control groups in three of the schools limits the conclusion. However, in those five schools in which adequate control groups were provided, THE WORD UP PROJECT students scored higher in three of the schools and scored the same in the other two despite the fact that in those schools the control group seemed to be a slightly higher academic achievement group.

In the three schools in which there was a strong prior academic achievement level for the control group, adjustments were made in the selection of the control group. Under those conditions, THE WORD UP PROJECT schools scored significantly higher in two of the three schools and in the third school the scores were even.

Raising state test scores is a difficult to achieve. These results suggest strongly that THE WORD UP PROJECT can play a significant role in raising state language arts/reading scores.