

**EVALUATION OF THE
WATERFORD EARLY READING PROGRAM
IN KINDERGARTEN
2005-06**

**STEPHEN POWERS, PH.D. AND CONNIE PRICE-JOHNSON, M.A.
CREATIVE RESEARCH ASSOCIATES
2030 E. BROADWAY, SUITE 221
TUCSON, ARIZONA 85719
(520) – 884-8667
SVPOWERS@AOL.COM**

JANUARY 2007

ACKNOWLEDGEMENTS

We would like to express our appreciation to the Tucson Unified School District's Office of Accountability and Research for their cooperation in retrieving relevant data from their database and for their expertise in merging and organizing the data for our analysis. This evaluation study would not be possible without their cooperation.

Stephen Powers, Ph.D. and Connie Price-Johnson, M.A.
Creative Research Associates, Inc.

ABSTRACT

Background

The Waterford Early Reading Program (WERP), a technology-based program for early elementary grades, was provided through Arizona all-day kindergarten funds to kindergarten students in 15 Title I elementary schools in the Tucson Unified School District (TUSD) in the 2005-06 school year. The purpose of this study is to evaluate the reading achievement of the kindergartners in the schools with the WERP and in a comparison group of 15 schools in the same district.

The schools where the WERP was implemented are identified in this report as Schools A-L. The comparison schools are identified as Schools M-AA.

Research Design

This evaluation design was a comparison-group study (quasi-experimental design) involving a treatment (WERP) implemented in 15 Title I schools ranked with the highest percentages of students on free/reduced lunch. A comparison group of 15 schools was selected from those with the next highest percentages of students on free/reduced lunch. The comparison schools did not receive the WERP.

Both matching techniques and statistical controls were used to make the groups similar in the analysis.

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Initial Sound Fluency, Letter Naming Fluency, Word Use Fluency, Phoneme Segmentation Fluency, and Nonsense Word Fluency and the district's Core Curriculum Standard Assessment (CCSA) reading test were given as pretests and posttests during the school year. In addition, the amount of time that each kindergartner used the WERP computer software was extracted from the software and used in the analysis.

Statistical Analysis

Dependent samples *t* tests were used to determine gains for the WERP and comparison groups, and gain score analysis was used to compare these gains for both groups. Analysis of covariance was used to adjust the posttest means for differences on the pretest means of the students.

Data were disaggregated by school, gender, ethnicity, pretest achievement quartiles, primary home language, and English language learner (ELL) status in order to determine patterns of achievement among these groups.

Major Findings

- The WERP kindergartners consistently outperformed the comparison group kindergartners on all outcome measures. Comparison school kindergartners did make substantial and in some cases outstanding gains from pretest to posttest. However, when these were compared with the gains of the WERP kindergartners, the WERP gains were substantially and significantly greater.
- Effect sizes of gains favored the WERP kindergartners, as well as effect sizes comparing the posttest achievement of the WERP kindergartners with the comparison kindergartners.
- WERP gains were greater for males in the WERP program than for males in the comparison group, and for females in the WERP than for females in the comparison group.
- WERP gains were greater for Whites, Hispanics, African Americans, Native Americans, and Asians than for their counterparts in the comparison group.
- WERP gains of White, African American, Hispanic, and Asian kindergartners were greater than the gains of White kindergartners in the comparison group.
- WERP gains of kindergartners with a primary home language of English, Spanish, and other languages were greater than their counterparts in the comparison group.
- WERP gains of kindergartners with a primary home language of Spanish were greater than the gains of English primary home language kindergartners in the comparison group. That is, WERP Spanish home language students who were learning English reading skills outperformed the comparison group English primary home language students.
- WERP gains of kindergartners in four different quartile levels of reading pretest achievement outperformed the comparison students with the largest gains in the top (fourth) quartile.
- WERP English language learners outperformed comparison group English language learners.
- WERP English language learners with emergent reading skills outperformed the non-English language learners (proficient English speakers) in the comparison group.
- Usage of the WERP software was found to be significantly correlated with the reading outcome measures and the pretest to posttest gains. It is an important finding that the greater the use of WERP content, the greater the reading gains.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	2
ABSTRACT	3
BACKGROUND AND PURPOSE	7
Background.....	7
Purpose.....	7
METHODS	8
Study Setting.....	8
Research Design.....	8
Study Population.....	8
Statistical Analyses	11
Measures of Outcomes.....	12
RESULTS	15
Effect Estimates of the Intervention	15
Intervention Effects on Subgroups.....	32
WERP Usage Effects	52
SUMMARY AND DISCUSSION	55
APPENDIX A: READING SCORES BY SCHOOL	57
APPENDIX B: READING PERCENTILES BY SCHOOL	58
APPENDIX C: READING SCORES BY SCHOOL for ALL STUDENTS	59
REFERENCES	61
ABOUT THE AUTHORS	63

TABLES

Table 1. Ethnicity of Kindergartners	9
Table 2. Kindergartners in the WERP Evaluation	10
Table 3. Kindergartners' Usage Minutes with the WERP.....	13
Table 4. Administration of DIBELS and CCSA Reading 2005-06.....	14
Table 5. Gains on All Outcome Measures (WERP Students With 1100 or More Usage Minutes)	16
Table 6. Gains on All Outcome Measures (All Students) .Error! Bookmark not defined.	
Table 7. Gains on All Outcome Measures (Students With 90 or More Days Attendance)	21
Table 8. Effect Size on All Outcome Measures (WERP Students With 1100 or More Usage Minutes)	23
Table 9. Effect Size on All Outcome Measures (All Students)	25
Table 10. Effect Size on All Outcome Measures (Students With 90 or More Days Attendance)	26
Table 11. ANCOVA and Effect Sizes on All Outcome Measures (WERP Students With 1100 or More Usage Minutes)	28
Table 12. ANCOVA and Effect Sizes on All Outcome Measures (All Students).....	30
Table 13. ANCOVA and Effect Sizes on All Outcome Measures (Students With 90 or More Days Attendance)	31
Table 14. WERP + Reading First and Comparison School Means and Gains	32
Table 15. WERP + Reading First and Comparison Gains on All Outcome Measures.....	34
Table 16. Males and Females on DIBELS Total Reading Score.....	36
Table 17. Ethnic Groups on DIBELS Total Reading Score	38
Table 18. Ethnic Groups on DIBELS Total Reading Score Grouped by Ethnicity.....	39
Table 19. Primary Home Languages on DIBELS Total Reading Score.....	43
Table 20. Four Achievement Quartiles on the DIBELS Total Reading Score	46
Table 21. ELL and Non-ELL Students on DIBELS Total Reading Score	49
Table 22. ANCOVA of WERP ELL and Comparison Non-ELL Students.....	51
Table 23. Correlations of WERP Usage, Achievement and Gains.....	52
Table 24. DIBELS Total Reading Score by WERP Usage Level	53
Table 25. DIBELS Total Reading Scores by School.....	57
Table 26. DIBELS Total Reading Local Percentiles by School.....	58
Table 27. All Students on DIBELS Total Reading Score.....	59
Table 28. Schools Ranked by Pretest Means on the DIBELS Total Reading Score	60

BACKGROUND AND PURPOSE

Background

The importance of early reading interventions has been argued by many researchers (Finn, Rotherham, & Hokanson, 2001; National Association for the Education of Young Children [NAEYC] & International Reading Association [IRA], 1998). Finn, et.al. (2001) has noted the problem of an achievement gap especially among ethnic groups, and how this gap widens as the years pass. The value of technology in the early grades and its integration with instruction has been noted by many (NAEYC, 1996). Walberg (2001), a well-known evaluator, reported after reviewing the Waterford Early Reading Program (WERP) that it was “spectacularly effective for beginning readers who initially scored in the lower third of the group” (p. 11).

The WERP is research-based and uses technology to teach young children to read, write, and keyboard. The program has three levels (Level 1, Level 2, and Level 3) and a separate Phonological Awareness component designed for K-2. Of particular interest to the Tucson Unified School District (TUSD) is that the Waterford Institute (2002a) has specified how the WERP addresses issues of the No Child Left Behind legislation in the major areas of emergent reading skills as well as the skills assessed by the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (Waterford, 2002b), which is used in TUSD to assess reading in kindergarten.

WERP was implemented in the kindergartens of 15 schools of TUSD in the 2005-06 school year. It is from this year of WERP implementation that the data for the present study comes.

Purpose

The purpose of this study was to evaluate the effectiveness of the WERP in the kindergartens of TUSD Title I schools and to compare the pretest-to-posttest reading achievement of the kindergartners during the 2005-06 school year with that of 15 comparison schools that did not receive the program.

METHODS

Study Setting

The TUSD school district is the largest school district in the Tucson area and the second largest in Arizona. It is a multiethnic school district with over 60,000 students, 3,700 teachers and 200 administrators.

Research Design

The present study used a non-equivalent group, quasi-experimental design. The treatment was the WERP software, which was loaded onto classroom computers in the 15 Title I schools with the greatest percentage of students on free/reduced lunch and the greatest percentage of students who were English language learners. The comparison schools were from the schools next in line to be eligible for Title I funds with the next highest percentages on free/reduced lunch.

Issues of missing data (Allison, 2001; Little & Rubin, 2002; McKnight, McKnight, & Figueredo, 2007) were considered, but the relationships among missing data and other variables were minor and therefore no substitutions or imputations were made.

Treatment

The treatment used was the WERP software, which was loaded onto classroom computers to supplement the district reading program. The WERP also has teacher manuals, videos, worksheets and other classroom and take-home materials. Students rotated in using the four to six computers in the classroom for 15 minutes at a time as recommended by the Waterford Institute. This study addresses implementation only to the extent that students used the WERP software.

Study Population

Treatment and Comparison Groups

WERP schools

Fifteen elementary schools were slated to receive the WERP. One of these opted instead for the Waterford Early Math & Science program, and in two additional schools it was impossible to extract the WERP usage data from the computers. The 12 remaining schools (740 kindergarten students) used the WERP software in the kindergartens as planned and usage data was available. These schools are identified in this report as Schools A-L. In two of these (Schools I and L), the low level of program usage resulted in their exclusion from most of the analyses.

Comparison Schools

The comparison group schools are identified in this report as Schools M-AA. 1480 kindergarten students participated in the comparison schools.

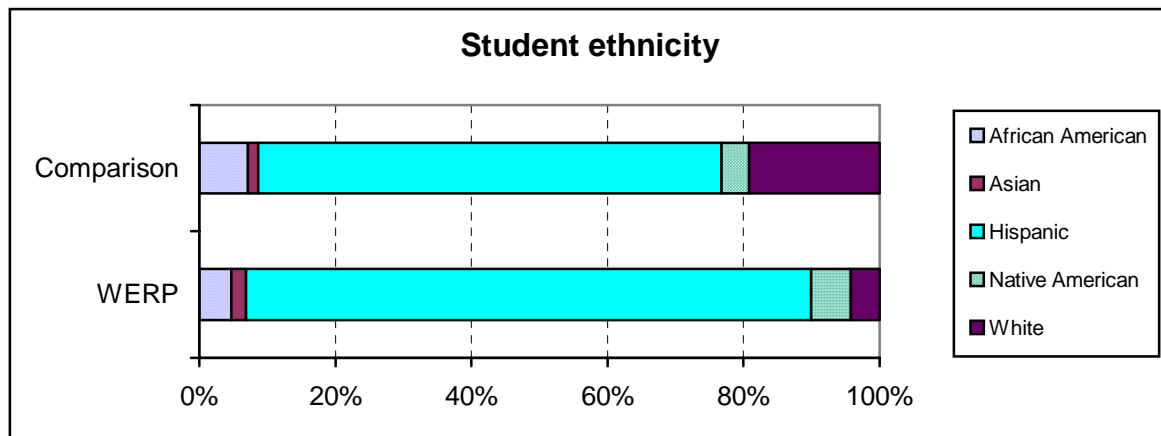
In the WERP group, 51% were males and 49% females. Of the comparison group, 50% were male and 50% were female. The primary language of the WERP group was 48% English, 49% Spanish and 3% other languages. In the comparison group, 68% used English as a primary language, 29% used Spanish, and 3% percent used another language. See Table1 for the ethnic composition of the WERP and comparison groups.

Table 1. Ethnicity of Kindergartners

Ethnicity	<u>WERP</u>		<u>Comparison</u>	
	N	%	N	%
African American	17	4.7	106	7.2
Asian	8	2.2	22	1.5
Hispanic	297	83.0	1008	68.1
Native American	21	5.9	60	4.1
White	15	4.2	284	19.2
Total	358	100.0	1480	100.0

Note. WERP students selected with 1100 minutes (6 months) or more usage of WERP.

Figure 1. Ethnicity of Kindergartners



Filtering of Students

Intent-to-treat analysis (ITT)

Research designs with randomization of treatment and control students often consider ITT analyses so that all students randomly assigned are entered into the analysis. This avoids the problem of the bias inherent in analyzing only students who are compliant with the research design (Ellenberg, 1996). Although students were not randomly

assigned, the ITT approach of assessing all students present in the program was followed along with other groupings. Efficacy subset analysis was also followed with those kindergarteners with 1100 minutes of WERP usage and with those with 90 days or more attendance.

The ITT group for the present study comprised all kindergarten students present at the beginning of the 2005-06 school year who were pretested. This distinction is important because in the public schools students enter classrooms throughout the school year and are not necessarily those who began the project at the beginning of the year. In the analyses, only those students with both pretest and posttest were included. This reduced the number of students in the study to 334 students in the WERP schools and 1211 in the comparison schools. See Table 2.

Filtered by 1100 minutes of WERP usage

Those students with at least 1100 minutes of usage of the WERP comprised the 1100-minutes group. This level of usage represents six months of using the program three times a week for 15 minutes at a time and was deemed sufficient to ensure an effect on student learning. This criterion excluded students from Schools I and L as mentioned above, reducing to 10 the number of schools and to 358 the number of WERP students in most of the analyses.

These WERP student gains were compared with those of the comparison group, which did not participate in the WERP and therefore had no minutes of usage.

Filtered by 90 Days Attendance

The gains of all of the WERP and comparison group kindergartners with 90 days or more attendance were compared. This criterion eliminated students from both groups who had poor attendance.

Table 2. Kindergartners in the WERP Evaluation

Group	Total	90 Days	1100 Mins	Pre-Posttest
WERP	740	636	358	334
Comparison	<u>1480</u>	<u>1183</u>	<u>1480*</u>	<u>1211</u>
Total	2220	1871	1838	1545

Note. *Students in Comparison group did not use Waterford, so number of students in the 1100 Mins column represents the nonWaterford students used in the comparative analyses.

Statistical Analyses

Whole-Group Analyses

Statistical analyses used included paired and independent samples *t* tests, gain score analysis, analysis of covariance, and effect size analysis. The students were grouped by 1) WERP students with at least 1100 minutes of program usage, 2) the intent-to-treat group (all students), and 3) students with at least 90 days attendance.

Gain Score Analysis

The pretest-to-posttest gains were compared using paired samples *t* tests to determine if the gains were statistically significant. In addition, the WERP and the comparison group gains were compared using independent samples *t* tests to determine if they were significantly different.

Analysis of Covariance

Analysis of covariance (ANCOVA) was used to adjust the posttest means of the WERP and the comparison groups for differences on the pretest. Following Winer (1971) and Kirk (1968), two of the issues of ANCOVA were examined: Significant differences between the WERP and comparison groups on the pretest would justify the use of ANCOVA to adjust for these differences. To apply ANCOVA, there should also be homogeneity of regressions of the groups being compared.

The WERP and comparison groups were compared to determine if their pretest means were significantly different. This was the case for Initial Sounds Fluency, Letter Naming Fluency, Word Use Fluency and Phoneme Segmentation Fluency. This suggests the use of ANCOVA for these outcome measures.

The WERP and comparison groups were compared to determine if there was homogeneity of regression. Only on Word Use Fluency and Phoneme Segmentation Fluency was there statistically significant heterogeneity of regression. Therefore, these results are presented with a caveat to the interpretation of the analyses of covariance.

Effect Size Analysis

Effect size analysis was completed to compare the WERP and comparison groups following the standard effect size methods (Cohen, 1977). The effect size that was considered useful was .20 for a small effect size, .50 for medium effect size, and .80 as a large effect size (Cohen, 1977).

Subgroup Analyses

In addition to the analyses of the WERP and comparison groups as a whole, several analyses of subgroups were carried out:

- The gains of three schools using the WERP and the Reading First programs were compared to those of three comparison schools with nearly the same pretest mean reading scores.
- Pretest-to-posttest gains of male and female kindergartners in the WERP and comparison schools were compared.
- Pretest-to-posttest gains of African-American, Asian, Hispanic, Native American and White kindergartners in the WERP and comparison schools were compared.
- Pretest-to-posttest gains of kindergartners with English, Spanish, and other primary home languages in the WERP and comparison schools were compared.
- Pretest-to-posttest gains of WERP and comparison group kindergartners in four reading achievement quartiles of the pretest were made in order to compare reading gains at different ability levels.
- Pretest-to-posttest gains of ELL kindergartners and non-ELL (English-proficient) kindergartners in the WERP and comparison groups were compared.

WERP Usage Effects

Correlational analyses between the total minutes of usage, reading achievement, and reading gains of the WERP students were completed to examine the relationship and effectiveness of the usage of WERP.

In addition, WERP students were categorized according to the number of minutes they used the WERP software. Their gains in reading were computed from pretest to posttest for each of the seven levels of usage.

Measures of Outcomes

Usage minutes

Staff of Pearson Digital Learning, which markets the Waterford Institute's products, collected the number of minutes each student used the WERP directly from the computers in the classrooms. TUSD's Office of Accountability and Research matched these records with student test scores and eliminated personal identifiers before the records were analyzed for this study.

Only students with sufficient exposure to the WERP (i.e., 1100 minutes or six months) were used in most of the analyses. Table 3 shows the number of students who used Level 1, Level 2 or the Phonological Awareness component of the WERP for any amount of time, and the range of minutes a single student spent on that level.

Table 3. Kindergartners' Usage Minutes with the WERP

Usage	N	Minutes
Reading Level 1: total minutes in course	700	0 – 2175
Reading Level 2: total minutes in course	203	0 – 2585
Phonological Awareness: total minutes	704	0 – 962
Total of all usage minutes	725	0 - 4003

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS), developed by researchers and specialists in early childhood education at the University of Oregon (Good & Kaminski, 2002), is a standardized assessment administered by TUSD to all kindergartners in the district three times a year and sent to the developers of the test for scoring. Scores are reported as raw scores and local percentiles. The DIBELS is composed of five subscales:

- Initial Sounds Fluency
- Letter Naming Fluency
- Word Use Fluency
- Phoneme Segmentation Fluency, and
- Nonsense Word Fluency scales.

Good and Kaminski (2002) reported psychometric research into the properties of the DIBELS. In summary, these authors report alternate-form and test-retest reliabilities and predictive and concurrent validities of the subscales to range from .36 to .91 with a median reliability of .66. The content of the subscales was carefully described and the constructs were described and related to the subscales so that one could conclude a high degree of content validity of these subscales. It was concluded that the DIBELS subscales were adequate for the present study.

The Waterford Institute (2002b) provided a detailed analysis of how the WERP activities were assessed by the DIBELS, as well as how the WERP addressed issues of the No Child Left Behind law (Waterford, 2002a).

For the purposes of the present study, the average of the five DIBELS subscales was computed to provide an overall measure of the pretest and posttest reading achievement, or Total Reading score, of the kindergarten students. The internal consistency (alpha) reliability of the test was .79. Only students who completed all administered subscales were entered into the Total Reading score.

Core Curriculum Standard Assessment (CCSA) Reading Test

The Core Curriculum Standard Assessment (CCSA) reading test was developed by TUSD for district use (TUSDStats, n.d.). It parallels the criterion-based Arizona's Instrument to Measure Standards (AIMS) and is given in the grades where the AIMS is not. The CCSA places kindergartners in four levels of achievement (0, 1, 2, 3), which

correspond to the AIMS levels of Falls Far Below, Approaches, Meets and Exceeds. Scores of 2 and 3 (Meets and Exceeds) are considered passing or mastery of the content. TUSD teachers gave the CCSA in the fall and spring, serving as a pretest and posttest along with the DIBELS.

Administration of reading measures in 2005-06

Table 4 shows the times during the school year when the DIBELS subtests and the CCSA were administered.

Table 4. Administration of DIBELS and CCSA Reading 2005-06

Measure	Fall 2005	Winter	Spring 2006
DIBELS			
Initial Sounds Fluency	X	X	
Letter Naming Fluency	X	X	X
Word Use Fluency	X	X	X
Phoneme Segmentation Fluency		X	X
Nonsense Word Fluency		X	X
Total Reading Score	X		X
CCSA Reading			
Reading Performance	X		X

RESULTS

Effect Estimates of the Intervention

Gain Score Analyses

The pretest-posttest scores were analyzed using paired samples *t* tests to determine if there were significant gains. In addition, the gains themselves were compared for the WERP and the comparison group using independent samples *t* tests. This was the most straightforward analysis of the gains of the WERP and comparison group.

Criticism of gain score analysis has focused on instances in which a low pretest score can give the appearance of great gains and great program impact, ignoring the tendency of regression toward the mean (Linn, 1981). To address some of these concerns, ANCOVA was used for additional analyses.

Findings

- WERP students with 1100 minutes or more use of WERP software outperformed comparison students on all reading outcome measures. The gains were substantial, statistically significant, and consistent.
- In the ITT group (all students), the WERP students outperformed the comparison students on all reading outcome measures. These gains were statistically significant and consistent, but were somewhat lower than the gains of the WERP students with 1100 usage minutes relative to the comparison group. In the ITT group, both WERP and comparison groups included all students without filtering for 1100 usage minutes of the WERP. This may account for the differences in the results of the two analyses.
- In the 90-day attendance group, the WERP group gains were greater than the comparison group gains, showing that the WERP students consistently and significantly outperformed the comparison students on all outcome reading measures.

Summary

The three separate analyses reported above were undertaken to determine the effectiveness of the WERP on student achievement when compared to the comparison group. The comparison group was of a higher socio-economic status as measured by percent of students on free/reduced lunch and had a higher percentage of English-proficient students.

In spite of these differences, the WERP students outperformed the comparison students consistently, significantly, and in many cases substantially in each of the analyses of 1) WERP students with 1100 usage minutes and the comparison group, 2) the ITT group with all students in the WERP and comparison groups, and 3) both WERP and comparison groups filtered for at least 90 days attendance during the 2005-06 school year.

Table 5. Gains on All Outcome Measures (WERP Students with 1100 or More Usage Minutes)

Measures	N	Pretest		Posttest		Gains	t	p
		M	SD	M	SD			
DIBELS: ISF								
WERP	334	4.87	5.71	23.50	13.88	18.63	24.87	.000
Comparison	1218	6.42	6.82	17.53	12.24	<u>11.11</u>	31.86	.000
WERP vs. Comparison						7.52***		
DIBELS: LNF								
WERP	334	4.44	8.17	43.47	16.32	39.03	44.92	.000
Comparison	1155	6.30	10.15	40.89	16.36	<u>34.59</u>	76.84	.000
WERP vs. Comparison						4.44***		
DIBELS: WUF								
WERP	325	3.57	7.15	32.93	20.53	29.36	26.85	.000
Comparison	998	4.94	10.40	32.34	20.87	<u>27.40</u>	40.55	.000
WERP vs. Comparison						1.96 ^a		
DIBELS: PSF								
WERP	355	21.05	15.93	46.43	15.05	25.38	30.72	.000
Comparison	1219	17.10	15.87	39.34	18.69	<u>22.24</u>	46.70	.000
WERP vs. Comparison						3.14**		
DIBELS: NWF								
WERP	355	18.26	14.61	38.67	20.59	20.41	24.81	.000
Comparison	1217	14.66	15.02	31.46	20.21	<u>16.80</u>	38.21	.000
WERP vs. Comparison						3.61***		
DIBELS: Total Reading								
WERP	334	10.62	6.98	33.59	12.42	22.97	48.80	.000
Comparison	1211	10.22	8.29	29.32	13.33	<u>19.10</u>	71.46	.000
WERP vs. Comparison						3.87***		
TUSD: CCSA Reading								
WERP	311	1.09	0.49	2.68	0.64	1.59	38.00	.000
Comparison	1263	1.07	0.59	2.41	1.02	<u>1.34</u>	46.44	.000
WERP vs. Comparison						0.25***		

Note. ISF = Initial Sounds Fluency, LNF = Letter Naming Fluency, WUF = Word Use Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency. *p < .05, ** p < .01, *** p < .001 from independent *t* tests comparing gains. ^ap = .142

Figure 2. Gains on All Outcome Measures

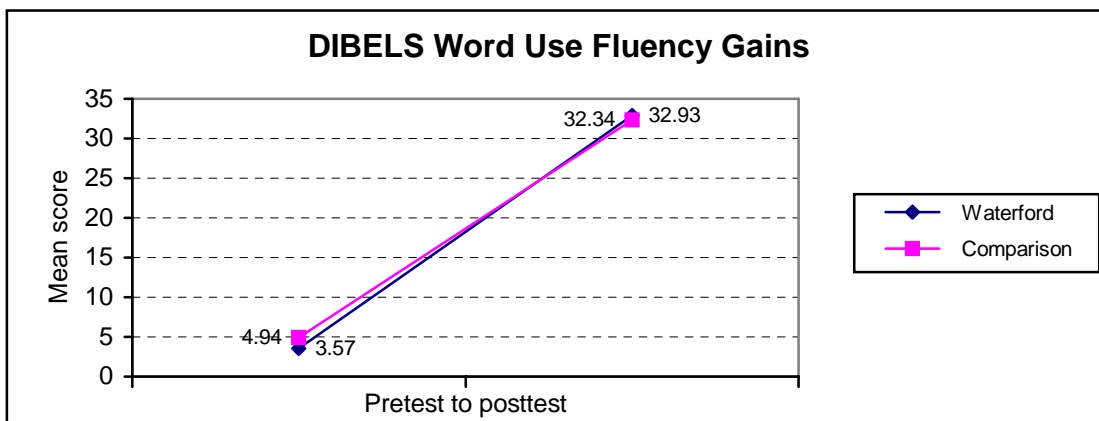
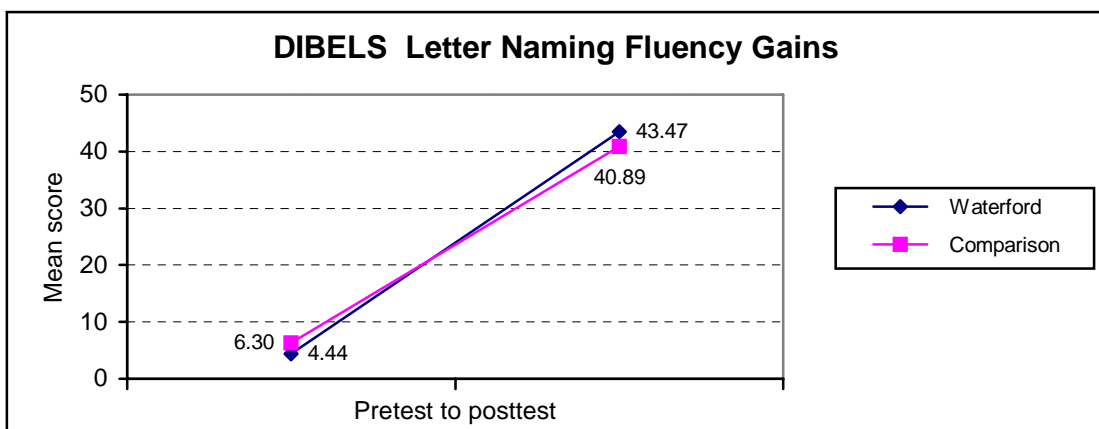
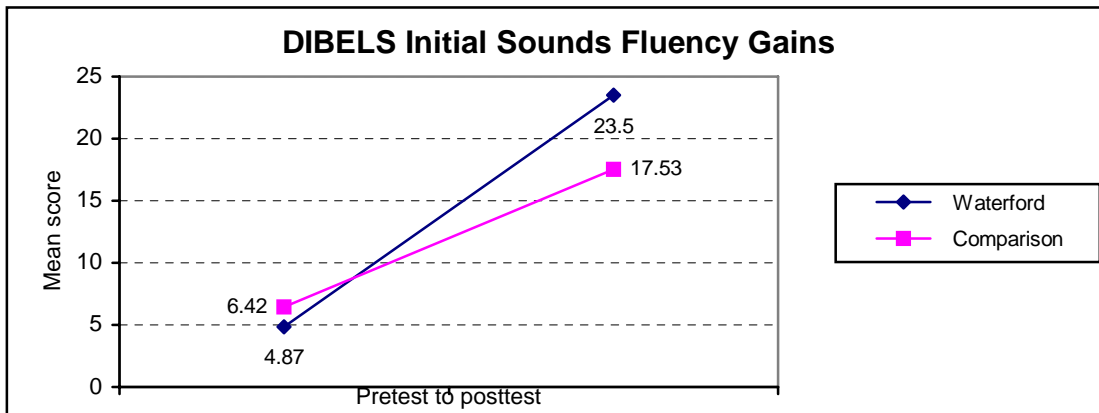


Figure 2. Gains on All Outcome Measures (continued)

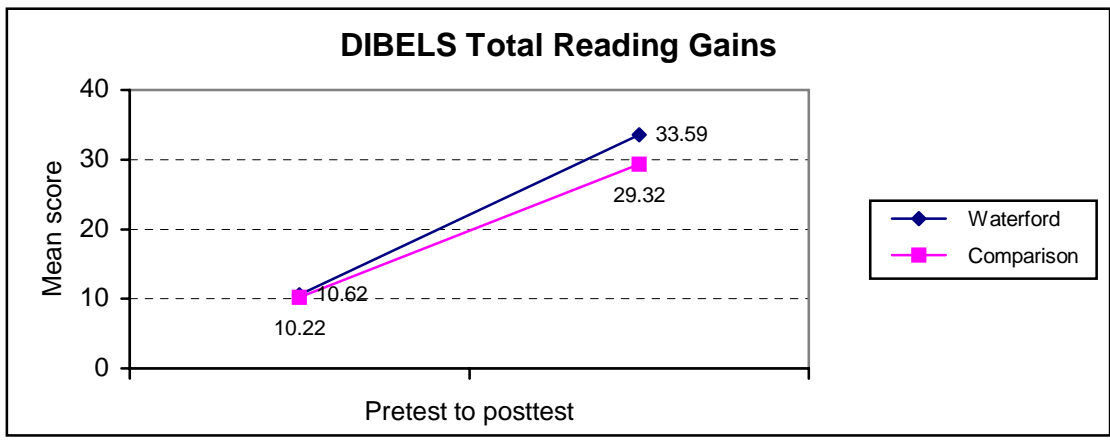
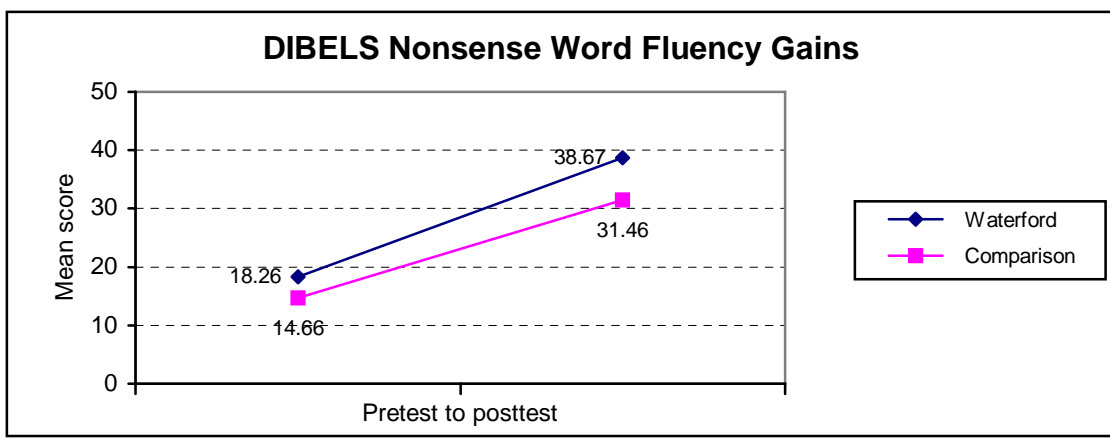
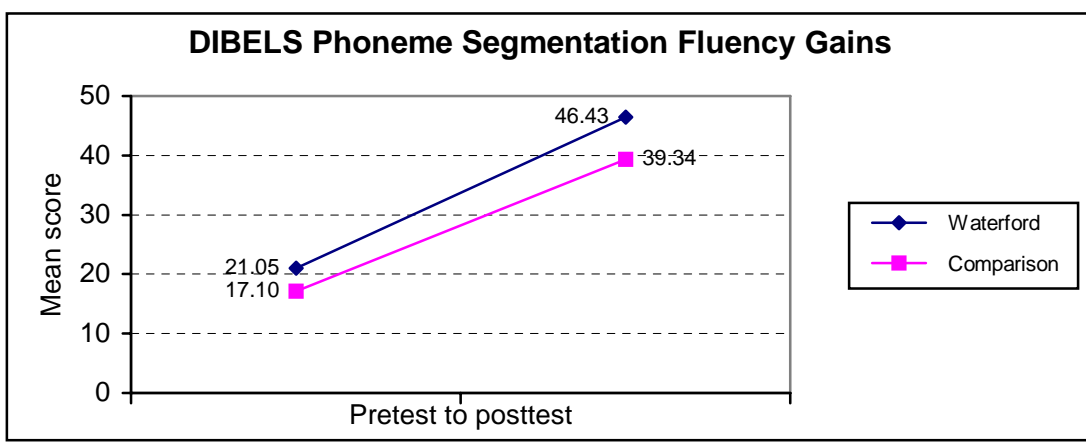


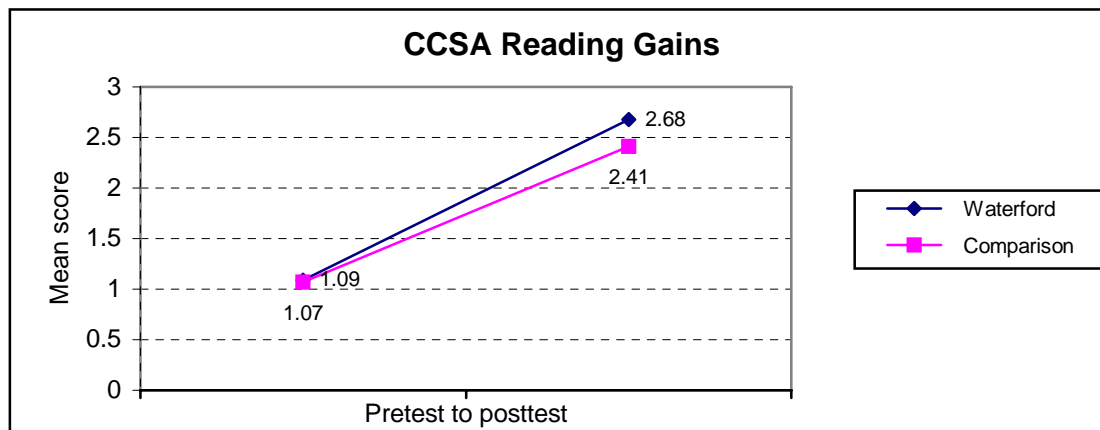
Figure 2. Gains on All Outcome Measures (continued)

Table 6. Gains on All Outcome Measures (All Students)

Measures	N	Pretest		Posttest		Gains	t	p
		M	SD	M	SD			
DIBELS: ISF								
WERP	640	5.26	5.99	20.81	13.41	15.55	29.25	.000
Comparison	1218	6.42	6.82	17.53	12.24	<u>11.11</u>	31.86	.000
WERP vs. Comparison						4.44***		
DIBELS: LNF								
WERP	630	4.02	7.39	41.99	17.15	37.97	58.67	.000
Comparison	1155	6.30	10.15	40.89	16.36	<u>34.59</u>	76.84	.000
WERP vs. Comparison						3.38***		
DIBELS: WUF								
WERP	584	3.73	8.18	31.29	19.82	27.56	34.72	.000
Comparison	998	4.94	10.40	32.34	20.87	<u>27.40</u>	40.55	.000
WERP vs. Comparison						.16***		
DIBELS: PSF								
WERP	670	20.44	16.52	44.43	16.78	23.99	38.83	.000
Comparison	1219	17.10	15.87	39.34	18.69	<u>22.24</u>	46.70	.000
WERP vs. Comparison						1.75*		
DIBELS: NWF								
WERP	670	15.66	14.10	34.59	20.01	18.93	33.52	.000
Comparison	1217	14.66	15.02	31.46	20.21	<u>16.80</u>	38.21	.000
WERP vs. Comparison						2.13**		
DIBELS: Total Reading								
WERP	636	10.08	7.34	31.11	12.58	21.03	62.16	.000
Comparison	1211	10.22	8.30	29.32	13.33	<u>19.10</u>	71.46	.000
WERP vs. Comparison						1.93***		
TUSD: CCSA Reading								
WERP	625	1.08	.48	2.62	.75	1.54	45.65	.000
Comparison	1263	1.07	.59	2.41	1.02	<u>1.34</u>	46.44	.000
WERP vs. Comparison						.20***		

Note. ISF = Initial Sounds Fluency, LNF = Letter Naming Fluency, WUF = Word Use Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency. * $p < .05$, ** $p < .01$, *** $p < .001$ from independent t tests comparing gains.

Table 7. Gains on All Outcome Measures (Students With 90 or More Days Attendance)

Measures	N	Pretest		Posttest		Gains	t	p
		M	SD	M	SD			
DIBELS: ISF								
WERP	640	5.26	5.99	20.81	13.41	15.55	29.25	.000
Comparison	1188	6.35	6.79	17.55	12.24	<u>11.20</u>	31.69	.000
WERP vs. Comparison						4.35***		
DIBELS: LNF								
WERP	628	4.03	7.40	42.09	17.07	38.06	58.95	.000
Comparison	1152	6.31	10.16	40.92	16.35	<u>34.61</u>	76.83	.000
WERP vs. Comparison						3.45***		
DIBELS: WUF								
WERP	582	3.75	8.19	31.31	19.80	27.56	34.72	.000
Comparison	995	4.95	10.41	32.36	20.86	<u>27.41</u>	40.50	.000
WERP vs. Comparison						.15***		
DIBELS: PSF								
WERP	661	20.50	16.54	44.38	16.85	23.88	38.51	.000
Comparison	1210	17.18	15.87	39.47	18.62	<u>22.29</u>	46.61	.000
WERP vs. Comparison						1.59*		
DIBELS: NWF								
WERP	661	15.76	14.13	34.68	20.08	18.92	33.21	.000
Comparison	1208	14.74	15.03	31.57	20.20	<u>16.83</u>	38.10	.000
WERP vs. Comparison						2.09**		
DIBELS: Total Reading								
WERP	636	10.08	7.34	31.11	12.58	21.03	62.16	.000
Comparison	1183	10.19	8.26	29.51	13.31	<u>19.32</u>	71.85	.000
WERP vs. Comparison						1.71***		
TUSD: CCSA Reading								
WERP	613	1.08	.48	2.65	.70	1.57	48.84	.000
Comparison	1197	1.07	.58	2.53	.90	<u>1.46</u>	56.68	.000
WERP vs. Comparison						.11**		

Note. ISF = Initial Sounds Fluency, LNF = Letter Naming Fluency, WUF = Word Use Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency. * $p < .05$, ** $p < .01$, *** $p < .001$ from independent t tests comparing gains.

Effect Size Analyses

Effect size analysis was used to compare the WERP and comparison groups following the standard effect size methods (Cohen, 1977). The effect size that was considered useful was .20 for a small effect size, .50 for medium effect size, and .80 as a large effect size (Cohen, 1977). The differences between the effect sizes and the ratio of the effect sizes were examined.

Findings

- WERP students with 1100 minutes or more use of WERP software outperformed the comparison students as measured by the effect sizes of their gains on all reading outcome measures. The differences in gains were substantial, consistent and larger than .20. Four of the effect sizes were small, one was medium, and the difference in effect sizes of the DIBELS Total Reading score and the CCSA reading test were medium (.62) to large (.85).
- In the ITT group (all students) the WERP student effect sizes were greater than the comparison effect sizes. The differences in effect sizes were small to moderate for four comparisons. The DIBELS Total Reading score and the CCSA reading test showed effect size differences of .42 and .53.
- In the 90-day attendance group, the WERP effect sizes were greater than the comparison effect sizes. The differences in effect sizes were small to moderate for all but three of the DIBELS reading scales. The DIBELS Total Reading score and the CCSA reading test showed effect size differences of .38 and .32.

Summary

The three separate analyses reported above were undertaken to determine the effectiveness of the WERP student achievement when compared to a comparison group.

The WERP students outperformed the comparison students consistently with greater effect sizes. Most of the effect size differences were at or above the criterion for a small effect size of .20.

Table 8. Effect Size on All Outcome Measures (WERP Students With 1100 or More Usage Minutes)

Measures	N	<u>Gains</u>		ES	Ratio
		M	GainSD		
DIBELS: ISF					
WERP	334	18.63	13.70	1.36	1.49
Comparison	1218	11.11	12.17	<u>.91</u>	
WERP vs Comparison				.45	
DIBELS: LNF					
WERP	334	39.03	15.88	2.46	1.09
Comparison	1155	34.59	15.30	<u>2.26</u>	
WERP vs Comparison				.20	
DIBELS: WUF					
WERP	325	29.36	19.72	1.49	1.16
Comparison	998	27.40	21.34	<u>1.28</u>	
WERP vs Comparison				.21	
DIBELS: PSF					
WERP	355	25.38	15.57	1.63	1.22
Comparison	1219	22.24	16.63	<u>1.34</u>	
WERP vs Comparison				.29	
DIBELS: NWF					
WERP	355	20.41	15.50	1.32	1.20
Comparison	1217	16.80	15.33	<u>1.10</u>	
WERP vs Comparison				.22	
DIBELS: Total Reading					
WERP	334	22.97	8.60	2.67	1.30
Comparison	1211	19.10	9.30	<u>2.05</u>	
WERP vs Comparison				.62	
TUSD: CCSA Reading					
WERP	311	1.59	.74	2.15	1.65
Comparison	1263	1.34	1.03	<u>1.30</u>	
WERP vs Comparison				.85	

Note. N = Number, M = Mean Gain, GainSD = Gain Standard Deviation, ES = Effect Size of Gain, Ratio = Ratio of WERP ES to comparison ES. The pretest-posttest effect size in the mean gain divided by the standard deviation (Walberg, 2001).

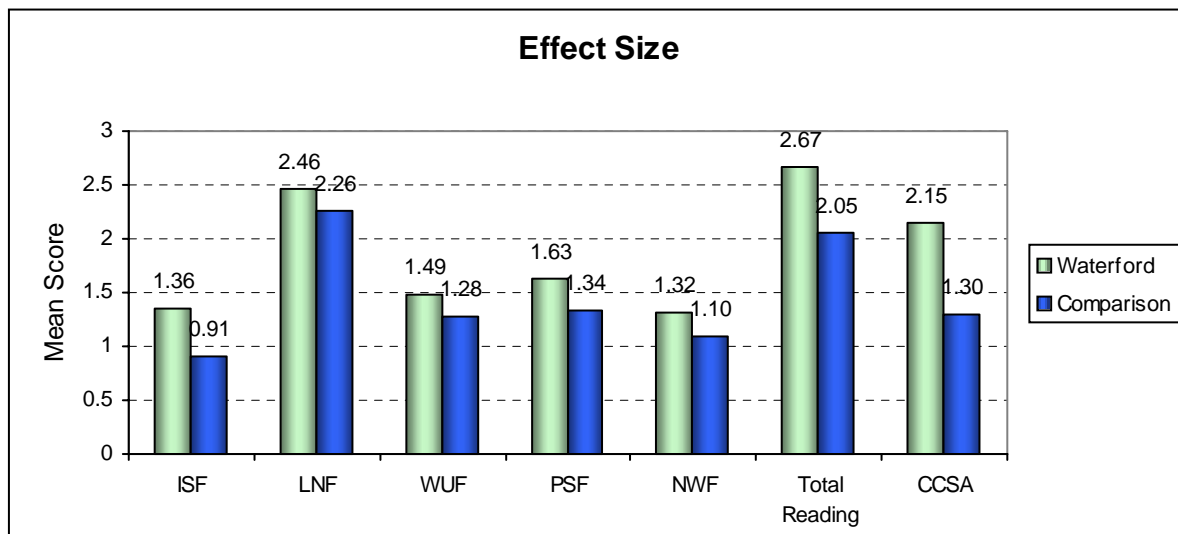
Figure 3. Effect Size on All Outcome Measures

Table 9. Effect Size on All Outcome Measures (All Students)

Measures	N	Gains		ES	Ratio
		M	GainSD		
DIBELS: ISF					
WERP	640	15.55	13.45	1.16	1.27
Comparison	1218	11.11	12.17	<u>.91</u>	
WERP vs Comparison				.25	
DIBELS: LNF					
WERP	630	37.97	16.24	2.33	1.03
Comparison	1155	34.59	15.30	<u>2.26</u>	
WERP vs Comparison				.07	
DIBELS: WUF					
WERP	584	27.56	19.18	1.44	1.13
Comparison	998	27.40	21.34	<u>1.28</u>	
WERP vs Comparison				.16	
DIBELS: PSF					
WERP	670	23.99	15.99	1.50	1.12
Comparison	1219	22.24	16.63	<u>1.34</u>	
WERP vs Comparison				.16	
DIBELS: NWF					
WERP	670	18.93	14.62	1.29	1.17
Comparison	1217	16.80	15.33	<u>1.10</u>	
WERP vs Comparison				.19	
DIBELS: Total Reading					
WERP	636	21.03	8.53	2.47	1.20
Comparison	1211	19.10	9.30	<u>2.05</u>	
WERP vs Comparison				.42	
TUSD: CCSA Reading					
WERP	625	1.54	.84	1.83	1.41
Comparison	1263	1.34	1.03	<u>1.30</u>	
WERP vs Comparison				.53	

Note. ISF = Initial Sounds Fluency, LNF = Letter Naming Fluency, WUF = Word Use Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency. N = Number, M = Mean Gain, PreSD = Pretest Standard Deviation, ES = Effect Size of Gain, Ratio = Ratio of WERP ES to comparison ES. The pretest-posttest effect size in the mean gain divided by the standard deviation (Walberg, 2001).

Table 10. Effect Size on All Outcome Measures (Students With 90 or More Days Attendance)

Measures	N	Gains		ES	Ratio
		M	GainSD		
DIBELS: ISF					
WERP	640	15.55	13.45	1.16	1.26
Comparison	1188	11.20	12.18	<u>.92</u>	
WERP vs Comparison				.24	
DIBELS: LNF					
WERP	628	38.06	16.18	2.35	1.04
Comparison	1152	34.61	15.29	<u>2.26</u>	
WERP vs Comparison				.09	
DIBELS: WUF					
WERP	582	27.56	19.16	1.44	1.13
Comparison	995	27.41	21.35	<u>1.28</u>	
WERP vs Comparison				.16	
DIBELS: PSF					
WERP	661	23.88	15.95	1.50	1.12
Comparison	1210	22.29	16.64	<u>1.34</u>	
WERP vs Comparison				.16	
DIBELS: NWF					
WERP	661	18.92	14.65	1.29	1.17
Comparison	1208	16.83	15.35	<u>1.10</u>	
WERP vs Comparison				.19	
DIBELS: Total Reading					
WERP	636	21.03	8.53	2.47	1.18
Comparison	1183	19.32	9.24	<u>2.09</u>	
WERP vs Comparison				.38	
TUSD: CCSA Reading					
WERP	613	1.57	.80	1.96	1.20
Comparison	1197	1.46	.89	<u>1.64</u>	
WERP vs Comparison				.32	

Note. ISF = Initial Sounds Fluency, LNF = Letter Naming Fluency, WUF = Word Use Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency. N = Number, M = Mean Gain, PreSD = Pretest Standard Deviation, ES = Effect Size of Gain, Ratio = Ratio of WERP ES to comparison ES. The pretest-posttest effect size in the mean gain divided by the standard deviation (Walberg, 2001).

Analysis of Covariance and Effect Size

Analysis of covariance (ANCOVA) was used to adjust the posttest means of the WERP and the comparison groups for differences on the pretest. Following this, the adjusted posttest means were compared as part of the ANCOVA with an *F* test, and effect sizes were calculated to determine if they were small (.20), medium (.50), or large (.80) using Cohen's (1977) criteria.

Findings

- WERP students with 1100 minutes or more using the software had significantly higher ($p < .001$) posttest mean scores than the comparison group except in Word Use Fluency. The effect sizes ranged from small to medium for all adjusted posttest mean comparisons except for Word Use Fluency.
- In the ITT group (all students), the WERP group had significantly higher ($p < .001$) posttest mean scores than the comparison group except in Word Use Fluency, where the posttest mean comparison was non-significant. The effect sizes ranged from small to medium for all adjusted posttest mean comparisons except for in Letter Naming Fluency, Word Use Fluency, and Nonsense Word Fluency.
- In the 90-day or more attendance group, the WERP group posttest means were significantly ($p < .01$) higher than the comparison group with the exception of Word Use Fluency. The effect sizes that were small ranged from .18 to .31. These effect sizes combined with the *F* tests indicated that the WERP program was effective in all areas but Word Use Fluency.

Summary

The three separate analyses reported above were undertaken to determine the effectiveness of the WERP on student achievement when compared to the comparison group with the posttest means adjusted for pretest differences.

For all three separate analyses, the WERP posttest means were higher than those of the comparison group except in Word Use Fluency.

For the three analyses with 21 ANCOVA *F* tests, 18 (86%) were significant favoring the WERP group. All of the comparisons of the WERP and comparison groups using the DIBELS Total Reading score as well as the CCSA reading resulted in statistically significant differences favoring the WERP group. In addition, effect sizes involving the Total Reading score ranged from .20 to .42, indicating an effect difference between the two groups.

Table 11. ANCOVA and Effect Sizes on All Outcome Measures (WERP Students With 1100 or More Usage Minutes)

Measures	N	Covariate		AdjPosttest		ES	F	p
		M	SD	M	SD			
DIBELS: ISF								
WERP	334	4.87	5.71	24.14	13.88	.56	81.57	.000
Comparison	1218	6.42	6.82	17.35	12.24			
DIBELS: LNF								
WERP	334	4.44	8.17	44.41	16.32	.25	16.33	.000
Comparison	1155	6.30	10.15	40.61	16.36			
DIBELS: WUF								
WERP	325	3.57	7.15	33.41	20.53	.06	.89	.345
Comparison	998	4.94	10.40	32.18	20.87			
DIBELS: PSF								
WERP	355	21.05	15.93	44.58	15.05	.31	26.22	.000
Comparison	1219	17.10	15.87	39.88	18.69			
DIBELS: NWF								
WERP	355	18.26	14.61	36.19	20.59	.26	18.65	.000
Comparison	1217	14.66	15.02	32.18	20.21			
DIBELS: Total Reading								
WERP	334	10.62	6.98	33.22	12.42	.42	46.16	.000
Comparison	1211	10.22	8.29	29.43	13.33			
TUSD: CCSA Reading								
WERP	311	1.09	0.49	2.67	0.64	.28	20.04	.000
Comparison	1263	1.07	0.59	2.41	1.02			

Note. ISF = Initial Sounds Fluency, LNF = Letter Naming Fluency, WUF = Word Use Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency. The effect size is the adjusted mean posttest difference divided by the square root of the ANCOVA mean squared residual.

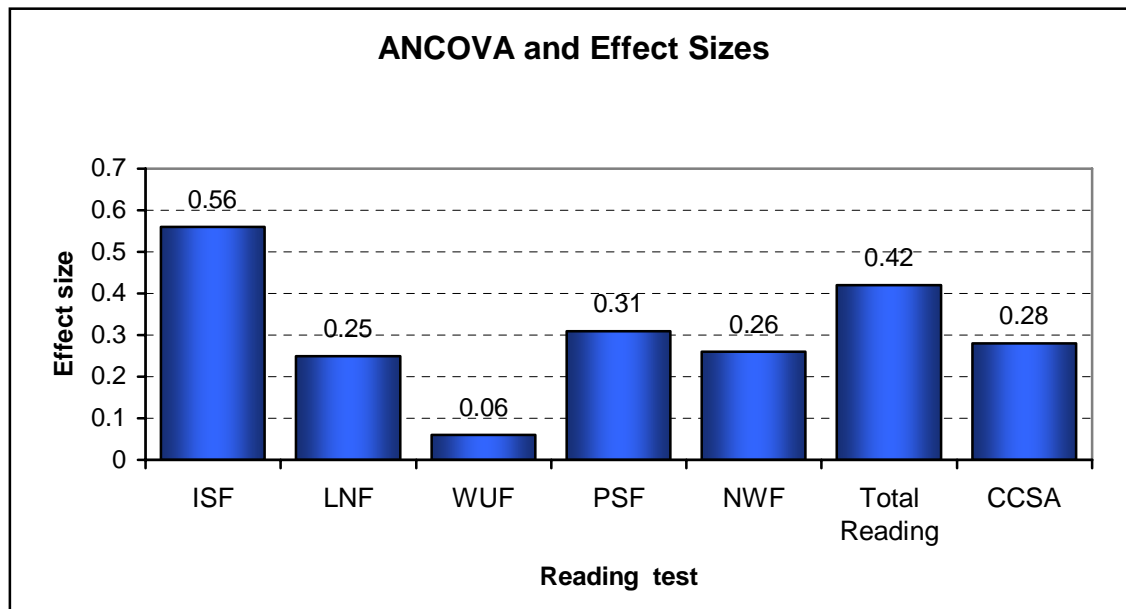
Figure 4. ANCOVA and Effect Sizes of on All Outcome Measures

Table 12. ANCOVA and Effect Sizes on All Outcome Measures (All Students)

Measures	N	Covariate		AdjPosttest		ES	F	P
		M	SD	M	SD			
DIBELS: ISF								
WERP	640	5.26	5.99	21.20	13.41	.32	41.96	.000
Comparison	1218	6.42	6.82	17.32	12.24			
DIBELS: LNF								
WERP	630	4.02	7.39	43.00	17.15	.17	12.13	.001
Comparison	1155	6.30	10.15	40.33	16.36			
DIBELS: WUF								
WERP	584	3.73	8.18	31.65	19.82	.00	.20	.651
Comparison	998	4.94	10.40	32.12	20.87			
DIBELS: PSF								
WERP	670	20.44	16.52	43.11	16.78	.20	17.41	.000
Comparison	1219	17.10	15.87	40.06	18.69			
DIBELS: NWF								
WERP	670	15.66	14.10	34.00	20.01	.15	9.44	.002
Comparison	1217	14.66	15.02	31.78	20.21			
DIBELS: Total Reading								
WERP	636	10.08	7.34	31.23	12.58	.22	20.23	.000
Comparison	1211	10.22	8.30	29.26	13.33			
TUSD: CCSA Reading								
WERP	625	1.08	.48	2.62	.75	.23	20.80	.000
Comparison	1263	1.07	.59	2.41	1.02			

Note. ISF = Initial Sounds Fluency, LNF = Letter Naming Fluency, WUF = Word Use Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency. The effect size is the adjusted mean posttest difference divided by the square root of the ANCOVA mean squared residual.

Table 13. ANCOVA and Effect Sizes on All Outcome Measures (Students With 90 or More Days Attendance)

Measures	N	Covariate		AdjPosttest		ES	F	P
		M	SD	M	SD			
DIBELS: ISF								
WERP	640	5.26	5.99	21.17	13.41	.31	40.30	.000
Comparison	1188	6.35	6.79	17.35	12.24			
DIBELS: LNF								
WERP	628	4.03	7.40	43.11	17.07	.18	12.77	.000
Comparison	1152	6.31	10.16	40.37	16.35			
DIBELS: WUF								
WERP	582	3.75	8.19	31.68	19.80	-.02	.20	.652
Comparison	995	4.95	10.41	32.15	20.86			
DIBELS: PSF								
WERP	661	20.50	16.54	43.08	16.85	.19	15.51	.000
Comparison	1210	17.18	15.87	40.18	18.62			
DIBELS: NWF								
WERP	661	15.76	14.13	34.09	20.08	.15	9.02	.003
Comparison	1208	14.74	15.03	31.90	20.20			
DIBELS: Total Reading								
WERP	636	10.08	7.34	31.20	12.58	.20	16.09	.000
Comparison	1183	10.19	8.26	29.46	13.31			
TUSD: CCSA reading								
WERP	613	1.08	.48	2.65	.70	.15	9.07	.003
Comparison	1197	1.06	.58	2.53	.90			

Note. ISF = Initial Sounds Fluency, LNF = Letter Naming Fluency, WUF = Word Use Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency. The effect size is the adjusted mean posttest difference divided by the square root of the ANCOVA mean squared residual. The covariate was the reading pretest.

Intervention Effects on Subgroups

This section examines the outcome measures when the student population is disaggregated by program, by gender, by ethnicity, by primary home language and by ELL status.

WERP and Reading First

The following pairs of schools were examined because these WERP schools had a high level of implementation of the program, were also Reading First schools, and were closely matched on the DIBELS pretest Total Reading Score with a comparison school that did not have either program. Thus School J (WERP) was matched with School X (comparison), School K (WERP) was matched with School V (comparison), and School H (WERP) was matched with School M (comparison).

The three schools with the WERP and the Reading First program outperformed the comparison schools with which they were matched, both together as a group and school by school. These schools were School J vs. School X; School K vs. School V; and School H vs. School M.

Analysis of the DIBELS subscales indicated that students receiving Reading First and the WERP performed significantly better than students in the comparison schools in Initial Word Fluency, Letter Naming Fluency, Nonsense Word Fluency, Total Reading score and in the CCSA reading assessment. In the DIBELS Total Reading score the difference was 4.90 points, statistically significant at the .001 level. See Tables 14 and 15 and Figures 5 and Figure 6.

Table 14. WERP + Reading First and Comparison School Means and Gains

Schools	N	Pretest	Posttest	Gain
WERP+ Reading First: School J	83	11.24	35.76	24.52
Comparison: School X	69	<u>11.33</u>	<u>32.63</u>	<u>21.30</u>
WERP vs Comparison		-.09	3.13	3.22
WERP+ Reading First: School K	79	11.89	37.99	26.10
Comparison: School V	97	<u>11.90</u>	<u>31.97</u>	<u>20.07</u>
WERP vs Comparison		-.01	6.02	6.03
WERP+ Reading First: School H	52	12.86	32.98	20.12
Comparison: School M	92	<u>13.06</u>	<u>29.56</u>	<u>16.50</u>
WERP vs Comparison		-.20	3.42	3.62

Note. WERP and comparison schools were matched by DIBELS Total Reading Score Pretest Mean.

Figure 5. WERP + Reading First and Comparison Schools Matched on DIBELS Reading Pretest Mean

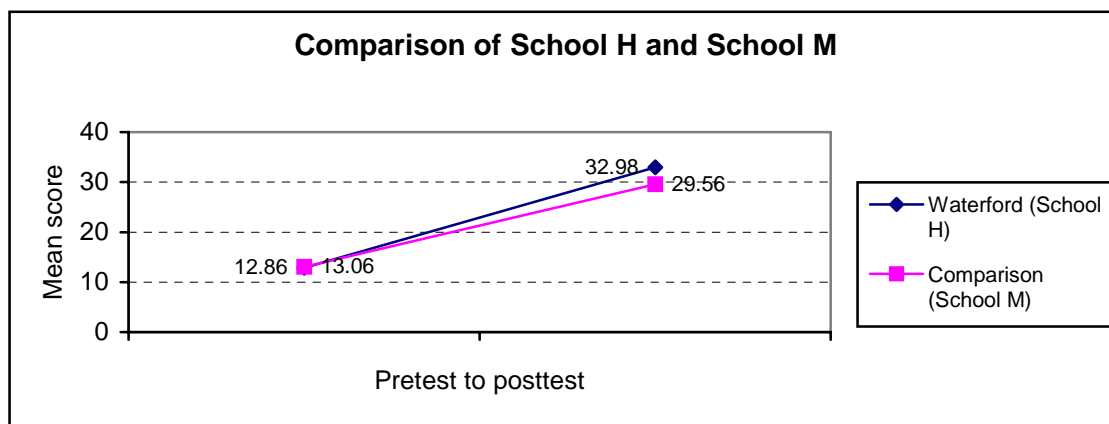
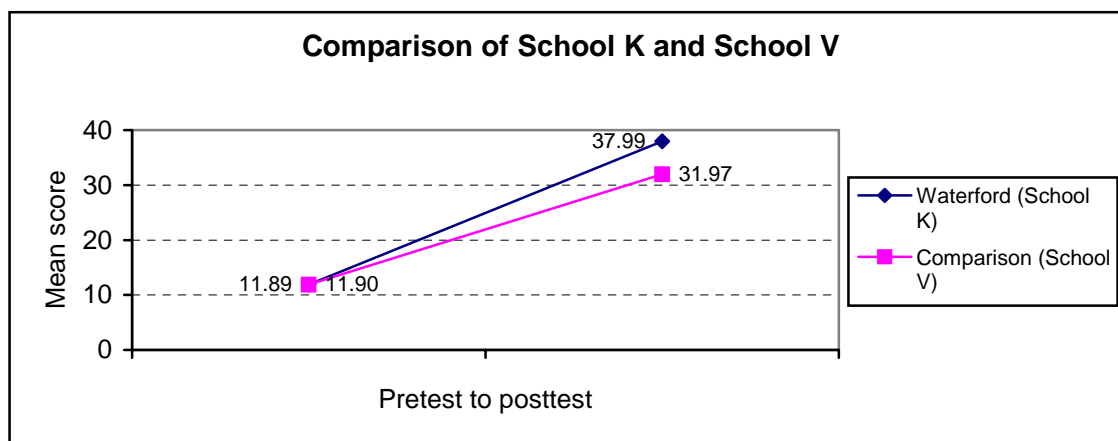
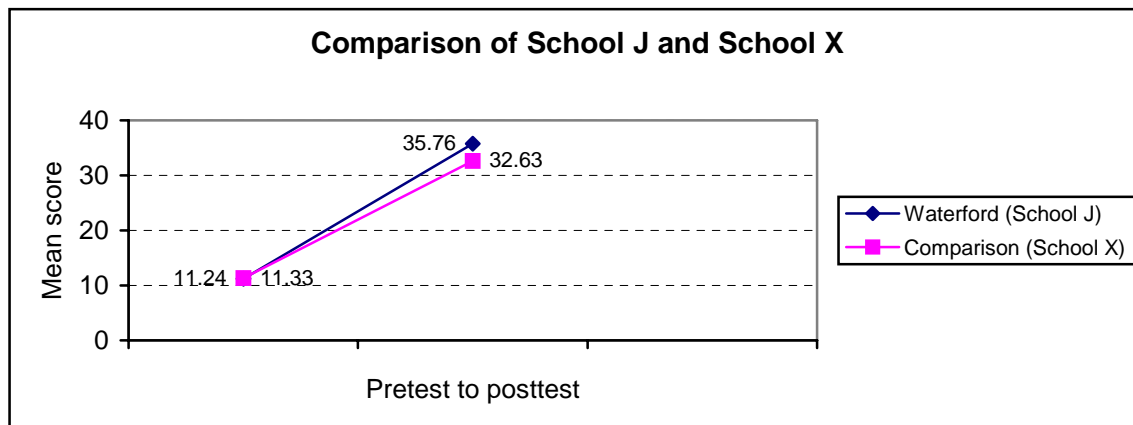
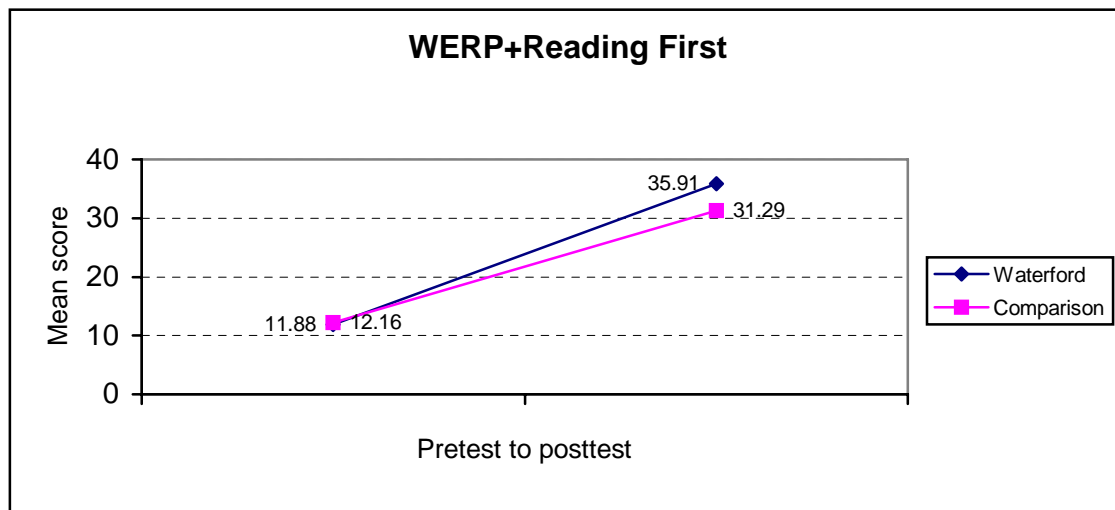


Table 15. WERP + Reading First and Comparison Gains on All Outcome Measures

Measures	N	Pretest		Posttest		Gain	t	p
		M	SD	M	SD			
DIBELS: ISF								
WERP	214	5.10	5.40	25.24	13.81	20.14	21.62	.000
Comparison	260	6.50	7.24	15.00	11.42	<u>8.50</u>	11.19	.000
WERP vs. Comparison						11.64***		
DIBELS: LNF								
WERP	214	4.88	8.99	45.95	15.81	41.07	38.81	.000
Comparison	239	5.04	9.67	41.79	16.55	<u>36.75</u>	36.34	.000
WERP vs. Comparison						4.32**		
DIBELS: WUF								
WERP	211	3.41	7.55	32.72	20.63	29.31	20.90	.000
Comparison	84	9.46	13.16	39.51	20.71	<u>30.05</u>	13.94	.000
WERP vs. Comparison						-0.74		
DIBELS: PSF								
WERP	230	23.96	16.16	50.53	11.11	26.57	25.81	.000
Comparison	255	19.37	16.32	44.93	18.58	<u>25.56</u>	24.12	.000
WERP vs. Comparison						1.01		
DIBELS: NWF								
WERP	230	20.52	14.02	43.94	20.12	23.42	22.33	.000
Comparison	253	17.17	13.96	34.91	19.01	<u>17.74</u>	18.85	.000
WERP vs. Comparison						5.68***		
DIBELS: Total Reading								
WERP	214	11.88	7.06	35.91	11.46	24.03	44.11	.000
Comparison	258	12.16	9.31	31.29	13.47	<u>19.13</u>	30.99	.000
WERP vs. Comparison						4.90***		
TUSD: CCSA Reading								
WERP	189	1.06	.50	2.72	.58	1.66	31.75	.000
Comparison	265	1.08	.60	2.36	1.07	<u>1.28</u>	19.84	.000
WERP vs. Comparison						0.38***		

Note. ISF = Initial Sounds Fluency, LNF = Letter Naming Fluency, WUF = Word Use Fluency, PSF = Phoneme Segmentation Fluency, NWF = Nonsense Word Fluency. WERP students selected with 1100 minutes (6 months) or more usage of WERP Reading Program. *p < .05, ** p < .01, *** p < .001 from independent *t* tests comparing gains.

Figure 6. WERP + Reading First and Comparison Gains on Total Reading Score

Gender

Females significantly outperformed males on the DIBELS Total Reading Score in both WERP and comparison schools. See Table 16.

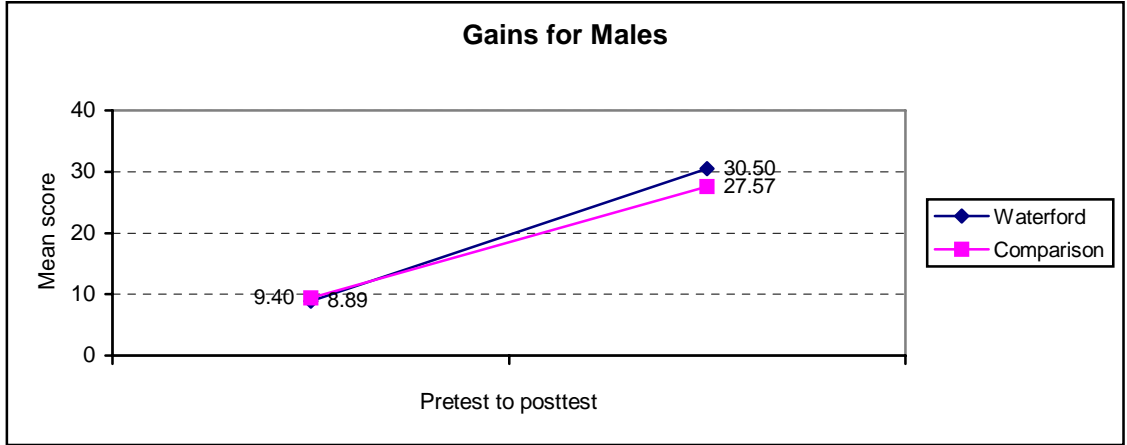
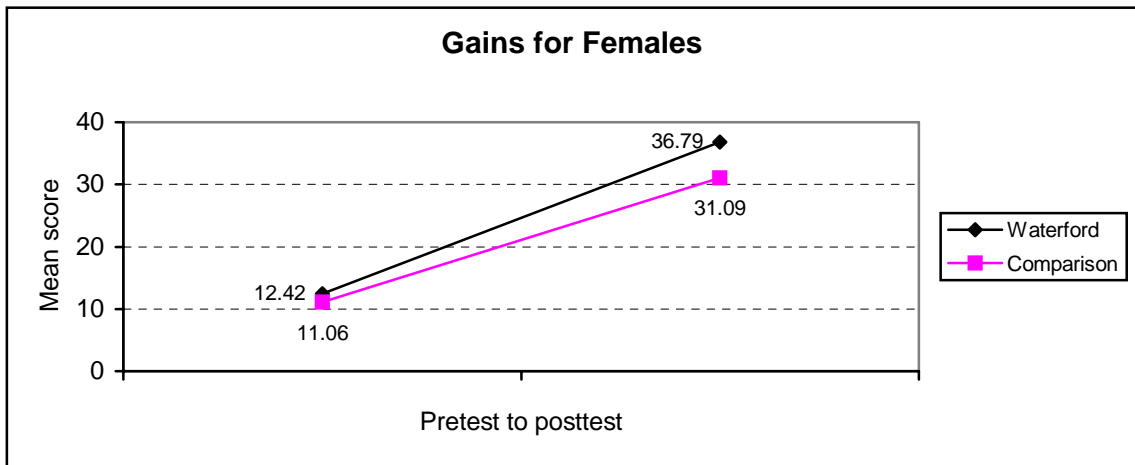
Males in the WERP schools outperformed males in comparison schools on the DIBELS Total Reading score, and WERP females outperformed comparison females. See Table 16 and Figure 7.

Table 16. Males and Females on DIBELS Total Reading Score

Group	N	Pretest		Posttest		Gain	t	p
		M	SD	M	SD			
WERP								
Female	164	12.42	7.17	36.79	11.67	24.37	38.07	.000
Male	170	8.89	6.33	30.50	12.38	<u>21.61</u>	32.09	.000
Female vs Male						2.76**		
Comparison								
Female	603	11.06	8.35	31.09	13.12	20.03	54.59	.000
Male	608	9.40	8.17	27.57	13.32	<u>18.17</u>	47.18	.000
Female vs Male						1.86***		
<hr/>								
Group	N	Pretest		Posttest		Gain	t	p
		M	SD	M	SD			
Male								
WERP	170	8.89	6.33	30.50	12.38	21.61	32.09	.000
Comparison	608	9.40	8.17	27.57	13.32	<u>18.17</u>	47.18	.000
WERP vs. Comparison						3.44***		
Female								
WERP	164	12.42	7.17	36.79	11.67	24.37	38.07	.000
Comparison	603	11.06	8.35	31.09	13.12	<u>20.03</u>	54.59	.000
WERP vs. Comparison						4.34***		

Note. WERP students selected with 1100 minutes (6 months) or more usage of WERP Reading Program.
*p < .05, ** p < .01, *** p < .001 from independent *t* tests comparing gains.

Figure 7. Males and Females on DIBELS Total Reading Score



Ethnicity

All ethnic groups, whether in the WERP schools or in the comparison schools, made important gains from pretest to posttest on the DIBELS Total Reading score. See Table 17.

A comparison of gains shows that all ethnic groups receiving the WERP made greater gains than their counterparts in the comparison group.

A surprising finding was that Hispanic (23.19), Asian (22.90), and African American (20.19) students in the WERP schools made greater gains pretest to posttest on the Total Reading score than did the White (19.82) students not receiving WERP.

The greatest gain pretest to posttest (26.23 points) was made by the White students in the WERP schools. These students also showed the greatest gain relative to their counterparts in the comparison schools for a statistically significant difference of 6.41 points.

The White WERP students showed the greatest gain relative to their comparison counterparts and the Native Americans showed the lowest. An analysis of variance (ANOVA) showed that the differences among ethnicities within the WERP group were not statistically significant.

Table 17. Ethnic Groups on DIBELS Total Reading Score

Group	N	Pretest		Posttest		Gain	t	p
		M	SD	M	SD			
WERP								
White	13	14.48	7.61	40.71	9.67	26.23	10.68	.000
African American	16	6.16	4.97	26.35	9.75	20.19	10.15	.000
Hispanic	279	10.65	6.91	33.84	12.48	23.19	45.45	.000
Native American	18	9.63	5.86	29.27	9.86	19.63	10.69	.000
Asian	8	14.55	9.53	37.45	16.60	22.90	5.32	.001
Comparison								
White	219	12.43	9.17	32.25	13.91	19.82	31.97	.000
African American	81	10.24	9.09	28.07	14.95	17.83	15.67	.000
Hispanic	850	9.59	7.92	28.65	12.95	19.06	60.33	.000
Native American	46	10.69	7.94	29.70	12.20	19.01	13.93	.000
Asian	15	12.31	7.56	30.47	15.95	18.16	6.18	.000

Note. WERP students selected with 1100 minutes (6 months) or more usage of WERP Reading Program. * $p < .05$, * $< .01$, *** $p < .001$ from independent t tests to compare gains.

Table 18. Ethnic Groups on DIBELS Total Reading Score Grouped by Ethnicity

Group	N	Pretest		Posttest		Gain	t	p
		M	SD	M	SD			
White								
WERP	13	14.48	7.61	40.71	9.67	26.23	10.68	.000
Comparison	219	12.43	9.17	32.25	13.91	<u>19.82</u>	31.97	.000
WERP vs. Comparison						6.41*		
African American								
WERP	16	6.16	4.97	26.35	9.75	20.19	10.15	.000
Comparison	81	10.24	9.09	28.07	14.95	<u>17.83</u>	15.67	.000
WERP vs. Comparison						2.36		
Hispanic								
WERP	279	10.65	6.91	33.84	12.48	23.19	45.45	.000
Comparison	850	9.59	7.92	28.65	12.95	<u>19.06</u>	60.33	.000
WERP vs. Comparison						4.13***		
Native American								
WERP	18	9.63	5.86	29.27	9.86	19.63	10.69	.000
Comparison	46	10.69	7.94	29.70	12.20	<u>19.01</u>	13.93	.000
WERP vs. Comparison						0.62		
Asian								
WERP	8	14.55	9.53	37.45	16.60	22.90	5.32	.001
Comparison	15	12.31	7.56	30.47	15.95	<u>18.16</u>	6.18	.000
WERP vs. Comparison						4.74		

Note. WERP students selected with 1100 minutes (6 months) or more usage of WERP Reading Program.

*p < .05, ** p < .01, *** p < .001 from independent t tests comparing gains.

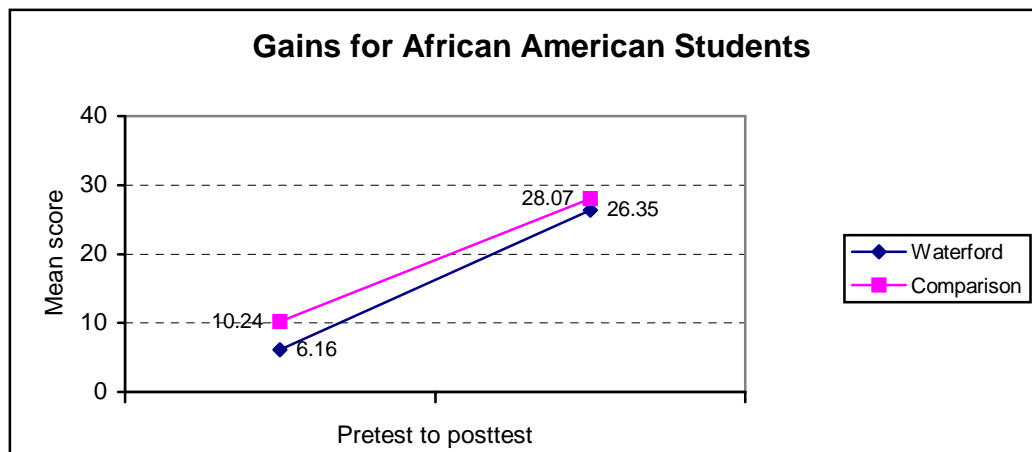
Figure 8. Ethnic Groups on DIBELS Total Reading Score

Figure 8. Ethnic Groups on DIBELS Total Reading Score (continued)

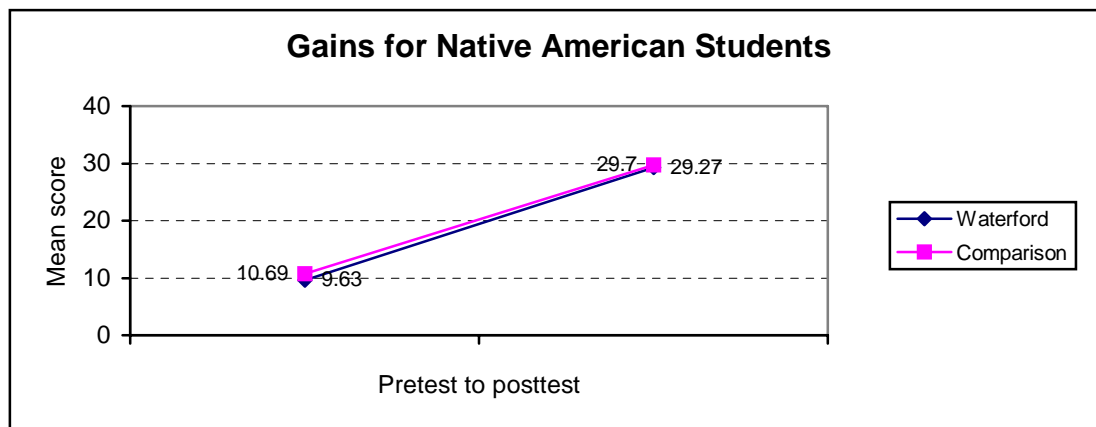
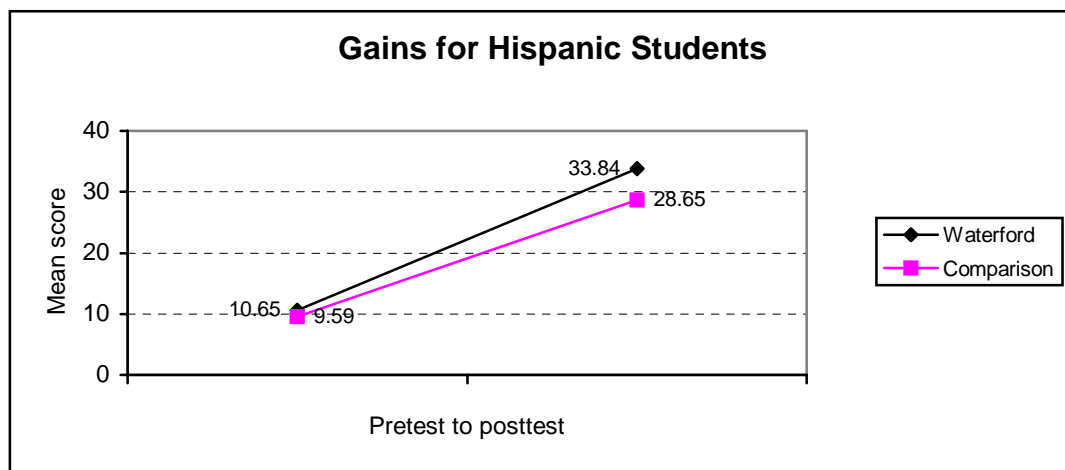
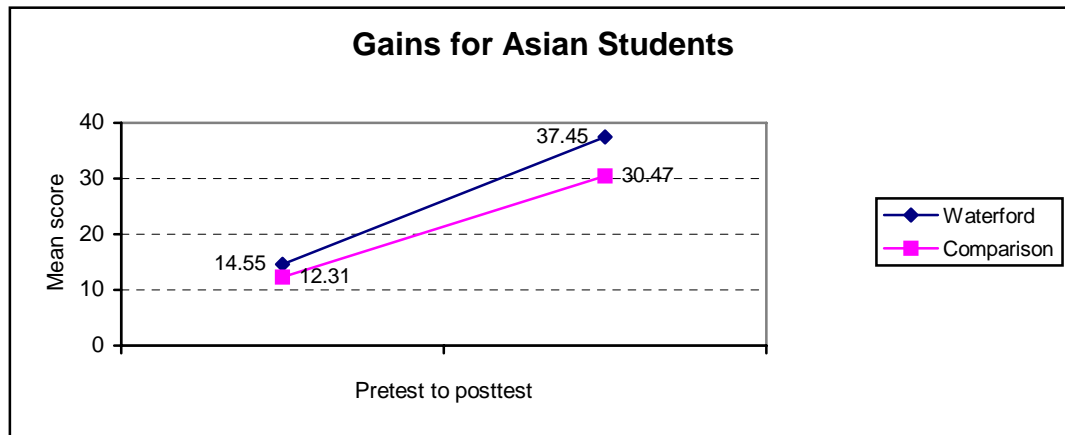


Figure 8. Ethnic Groups on DIBELS Total Reading Score (continued)

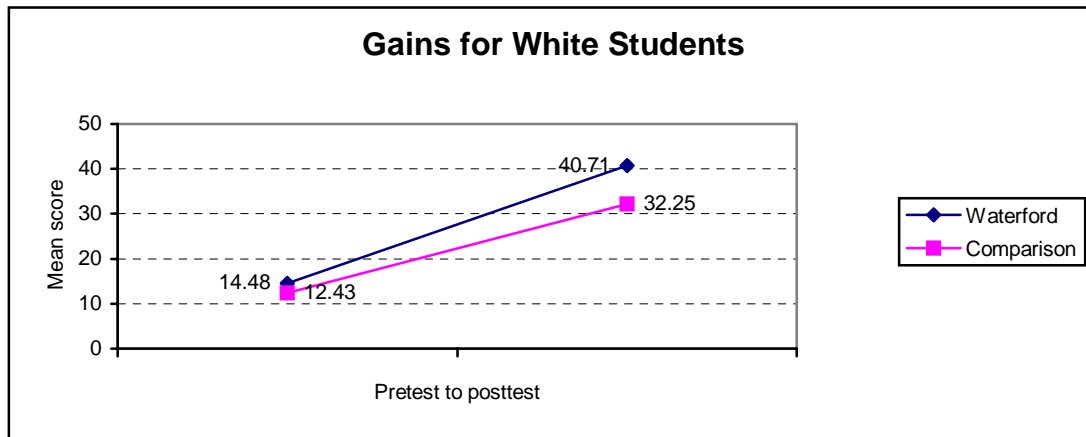
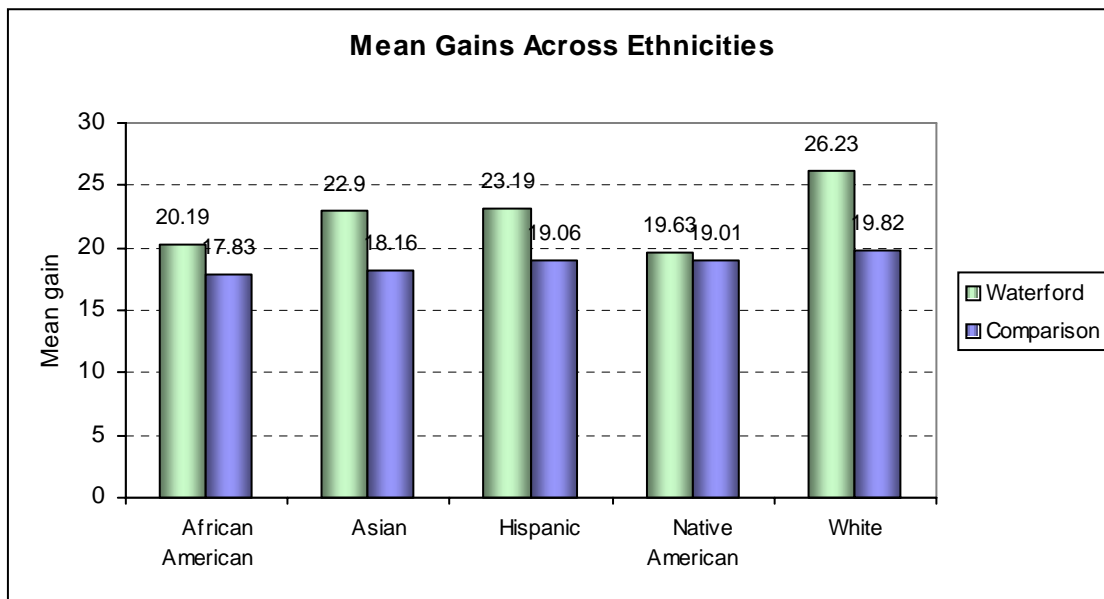


Figure 9. Mean Gains by Ethnicity on DIBELS Total Reading Score



Primary Home Language

Whether their primary home language was English, Spanish or another language, WERP students outperformed their counterparts in the comparison group on the DIBELS Total Reading Score. This difference was statistically significant for the English and the Spanish home language groups.

WERP students with Spanish (22.21) as their primary home language significantly outperformed in gains the comparison group students who spoke English as their primary home language (20.15).

The greatest gain in pretest to posttest scores was by the English-speaking WERP students, who gained 24.18 points.

The WERP group with the greatest gain (5.56 points) relative to the comparison group was that of students who spoke a primary home language other than English or Spanish. This diverse group includes refugee children who often have a history of upheavals, trauma and no prior school experience.

Table 19. Primary Home Languages on DIBELS Total Reading Score

Group	N	Pretest		Posttest		Gain	t	p
		M	SD	M	SD			
WERP								
English	160	11.41	6.69	35.58	11.66	24.18	36.68	.000
Spanish	163	10.08	7.13	32.29	12.84	22.21	32.52	.000
Other	11	7.11	7.48	23.82	10.62	16.71	7.79	.000
Comparison								
English	823	11.42	8.77	31.57	13.17	20.15	62.79	.000
Spanish	362	7.94	6.53	25.23	12.17	17.29	36.49	.000
Other	26	4.15	4.88	15.30	12.31	11.15	6.43	.000
Group	N	Pretest		Posttest		Gain	t	p
		M	SD	M	SD			
English								
WERP	160	11.41	6.69	35.58	11.66	24.18	36.68	.000
Comparison	823	11.42	8.77	31.57	13.17	<u>20.15</u>	62.79	.000
WERP vs. Comparison						4.03***		
Spanish								
WERP	163	10.08	7.13	32.29	12.84	22.21	32.52	.000
Comparison	362	7.94	6.53	25.23	12.17	<u>17.29</u>	36.49	.000
WERP vs. Comparison						4.92***		
Other								
WERP	11	7.11	7.48	23.82	10.62	16.71	7.79	.000
Comparison	26	4.15	4.88	15.30	12.31	<u>11.15</u>	6.43	.000
WERP vs. Comparison						5.56		

Note. Other languages are Af-Mayma, Amharic, Arabic, Cantonese, Persian, Filipino, French, Laotian, Marshallese, Portuguese, Russian, Somali, and Vietnamese. WERP students selected with 1100 minutes (6 months) or more usage of WERP Reading Program.

* $p < .05$, ** $p < .01$, *** $p < .001$ from independent t test comparing gains.

Figure 10. Primary Home Languages on DIBELS Total Reading Score

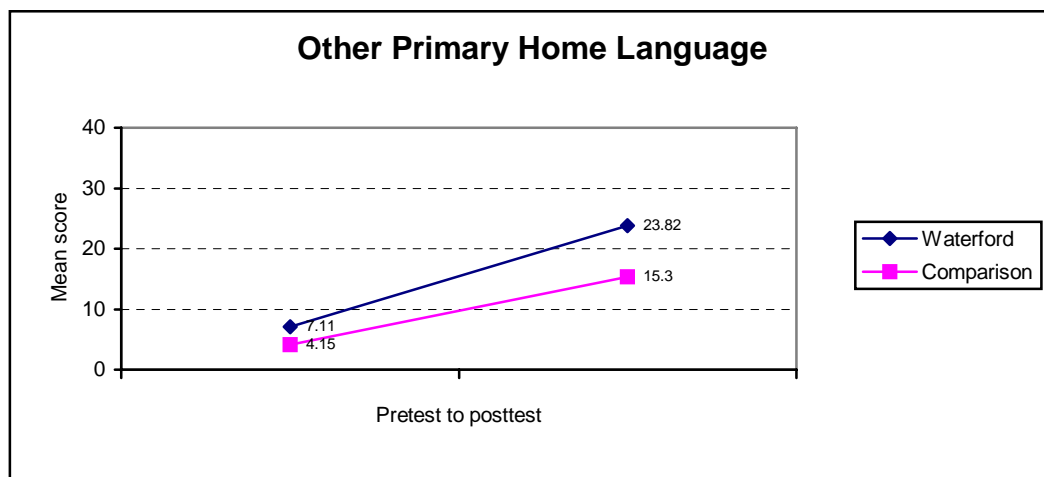
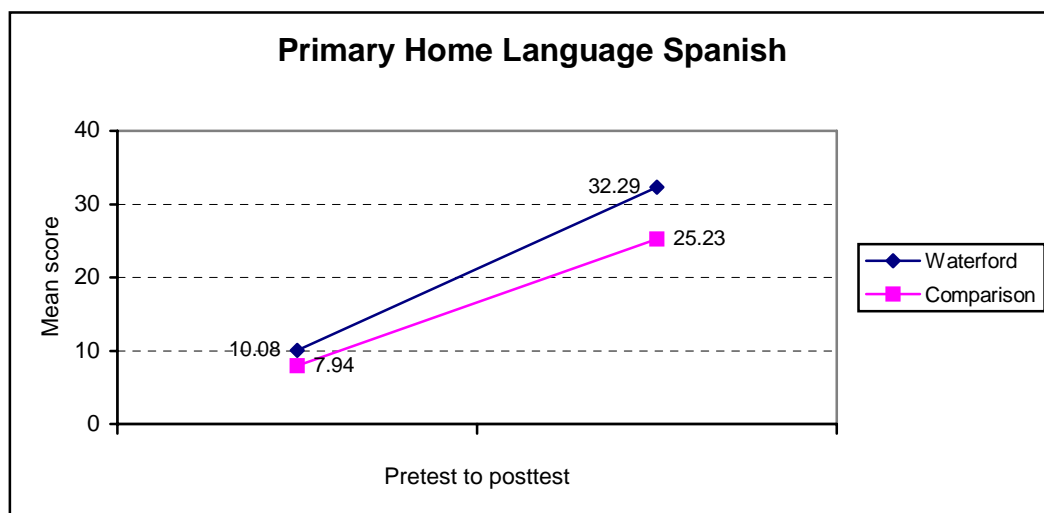
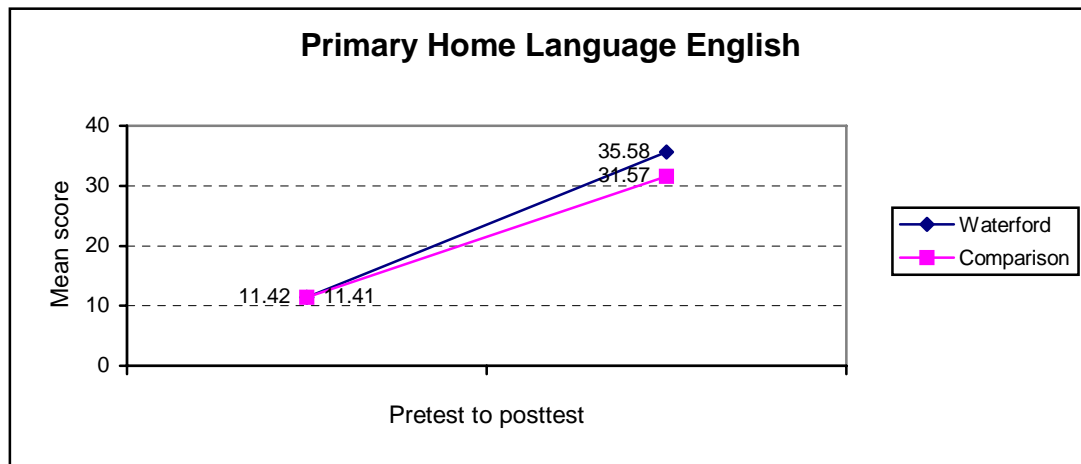
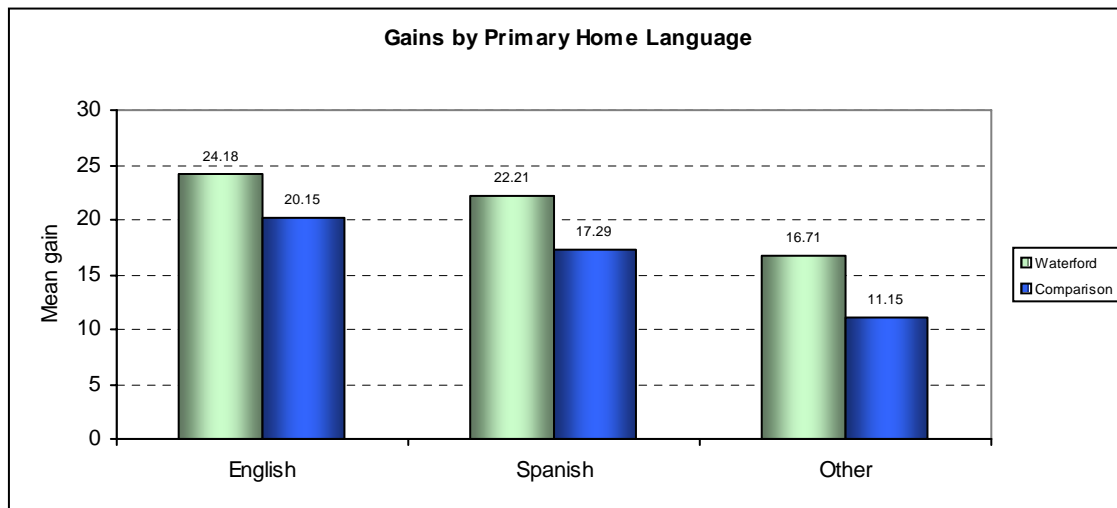


Figure 11. Gains by Primary Home Language

Achievement Quartile

Kindergartners were grouped based on the DIBELS Total Reading Score pretest into four quartiles. Students in each quartile at the WERP schools scored higher on the posttest than students in the same quartile at the comparison schools.

When the pretest–posttest gains were compared, WERP kindergartners in the 2nd, 3rd, and 4th quartile made significantly greater gains than their counterparts in the 2nd, 3rd and 4th quartiles. The greatest gain (25.20 points) was made by the 4th quartile students in the WERP schools. This result suggests that the WERP provides content to allow children who come with more preliteracy experience to make more rapid gains. See Table 20.

Table 20. Four Achievement Quartiles on the DIBELS Total Reading Score

Group	N	Means		Gain	t	p
		Pretest	Posttest			
<u>1st Quartile</u>						
WERP	60	1.60	18.24	16.64	13.64	.000
Comparison	304	1.66	16.35	<u>14.69</u>	27.62	.000
WERP vs Comparison				1.95		
<u>2nd Quartile</u>						
WERP	75	6.00	29.70	23.70	27.42	.000
Comparison	316	6.12	26.12	<u>20.00</u>	41.49	.000
WERP vs Comparison				3.70**		
<u>3rd Quartile</u>						
WERP	113	11.63	35.77	24.14	34.23	.000
Comparison	278	11.36	32.59	<u>21.23</u>	40.91	.000
WERP vs Comparison				2.91**		
<u>4th Quartile</u>						
WERP	86	19.62	44.82	25.20	28.08	.000
Comparison	313	21.68	42.26	<u>20.58</u>	39.40	.000
				4.62***		

Note. WERP students selected with 1100 minutes (6 months) or more usage of WERP Reading Program. The four achievement quartiles (25%) based on all students' rankings on the DIBELS Total Pretest Score. * $p < .05$, ** $p < .01$, *** $p < .001$ from independent t tests comparing gains.

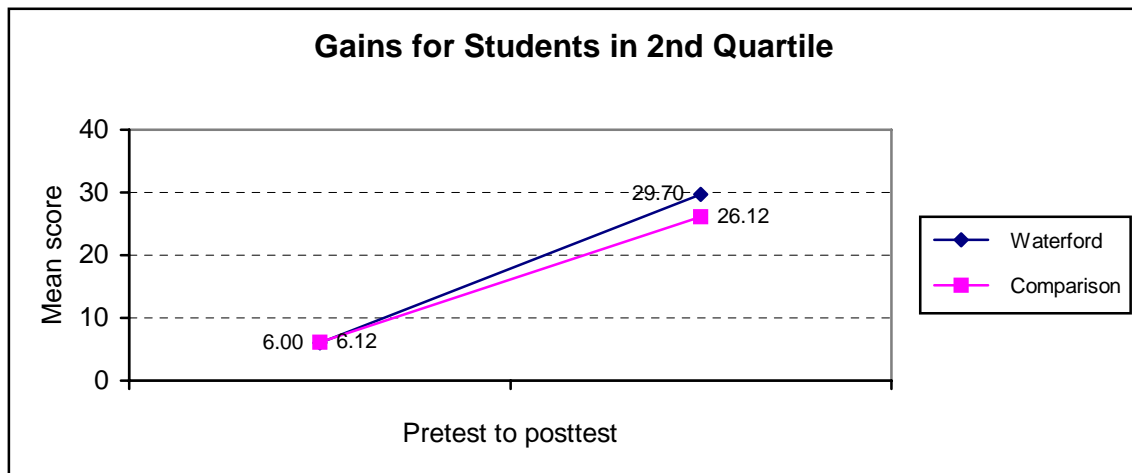
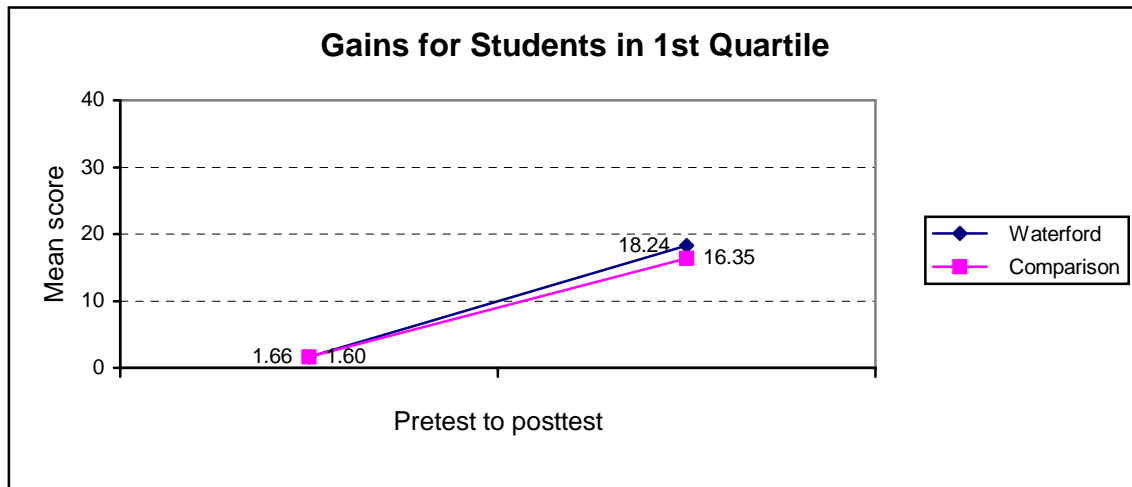
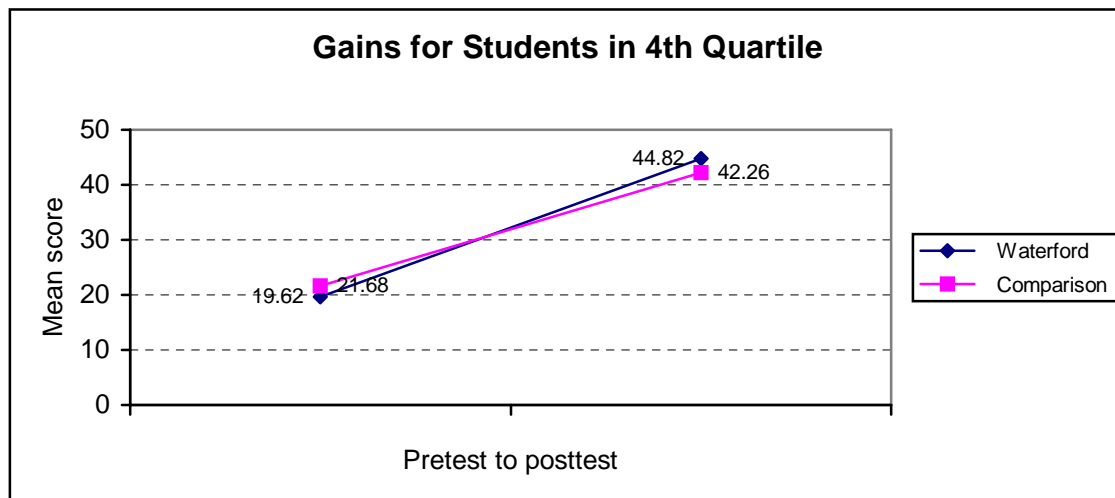
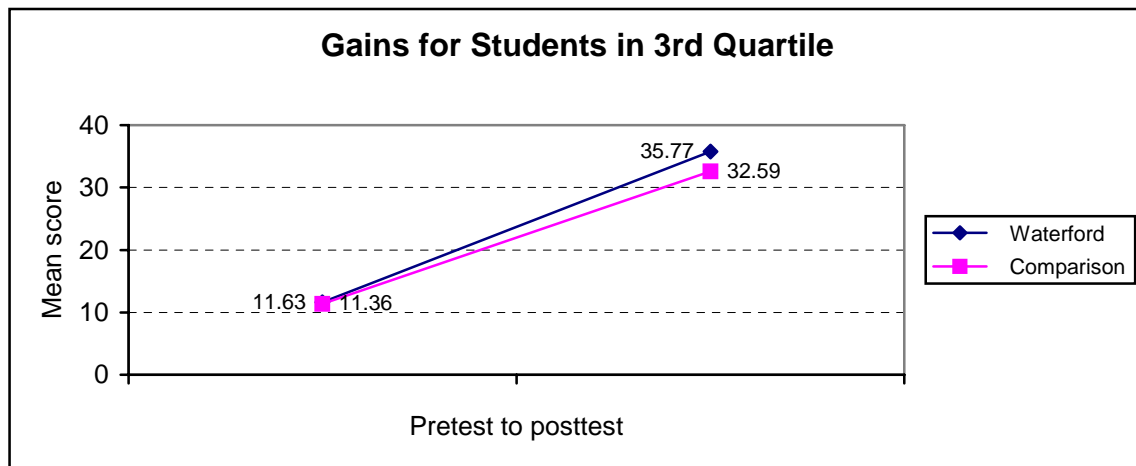
Figure 12. Gains by Achievement Quartile

Figure 12. Gains by Achievement Quartile (continued)

English Language Learner Status

Both ELL students and non-ELL (English-proficient) students made significantly greater gains pretest to posttest than their counterparts in the comparison group. That is, the WERP ELL students significantly outperformed comparison ELL students, and non-ELL (English-proficient) WERP students significantly outperformed the non-ELL comparison students.

WERP ELL students' gain (24.48) was compared to the comparison non-ELL (English-proficient) students' gain (19.67) using ANCOVA in order to adjust for initial differences. The WERP ELL students statistically outperformed the English-speaking students in the comparison group in gains. See Table 22.

It was unusual to find that students who did not speak English well (ELL students) outperformed students who were English-proficient (non-ELL students) in learning to read English.

Table 21. ELL and Non-ELL Students on DIBELS Total Reading Score

Measures	N	Pretest		Posttest		Gain	t	p
		M	SD	M	SD			
ELL students								
WERP	164	9.12	6.65	31.23	12.95	22.11	32.37	.000
Comparison	329	6.38	5.66	22.77	12.04	<u>16.39</u>	32.43	.000
WERP vs. Comparison						5.72***		
Non-ELL students								
WERP	170	12.07	7.00	35.86	11.49	23.79	36.94	.000
Comparison	882	11.66	8.66	31.77	12.97	<u>20.11</u>	65.26	.000
WERP vs. Comparison						3.68***		

Note. WERP students selected with 1100 minutes (6 months) or more usage of WERP Reading Program.
*p < .05, ** p < .01, *** p < .001 from independent *t* tests comparing gains.

Figure 13. Pretest to Posttest Scores by English Language Learner Status

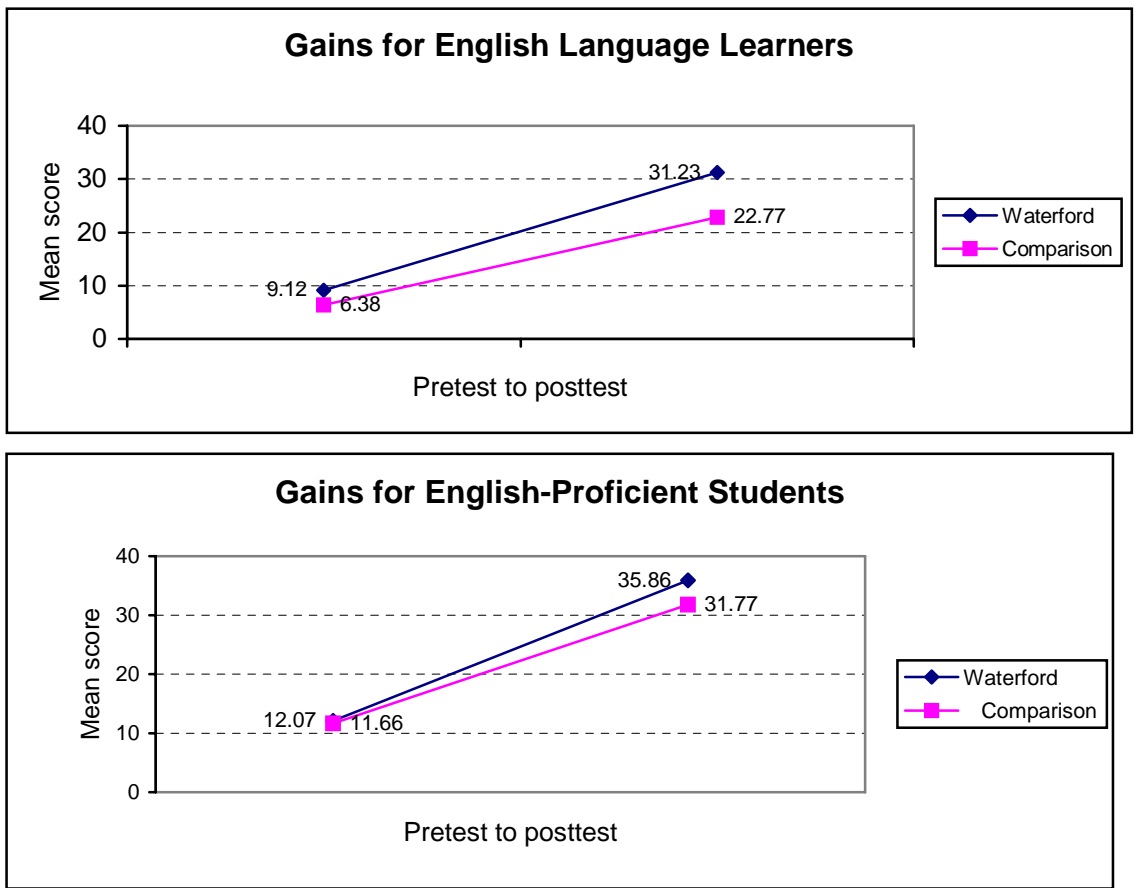


Figure 14. Gains by Language Learning Status

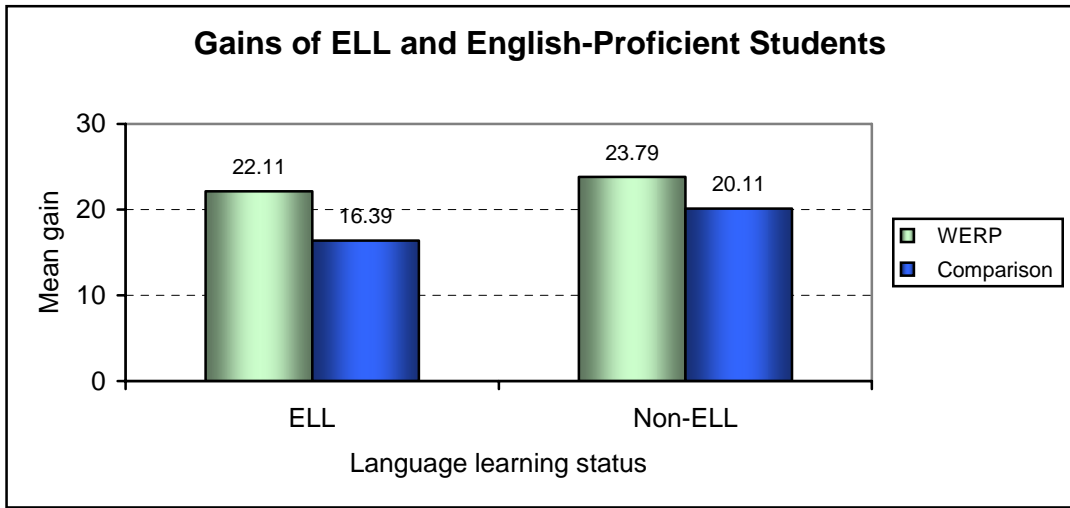
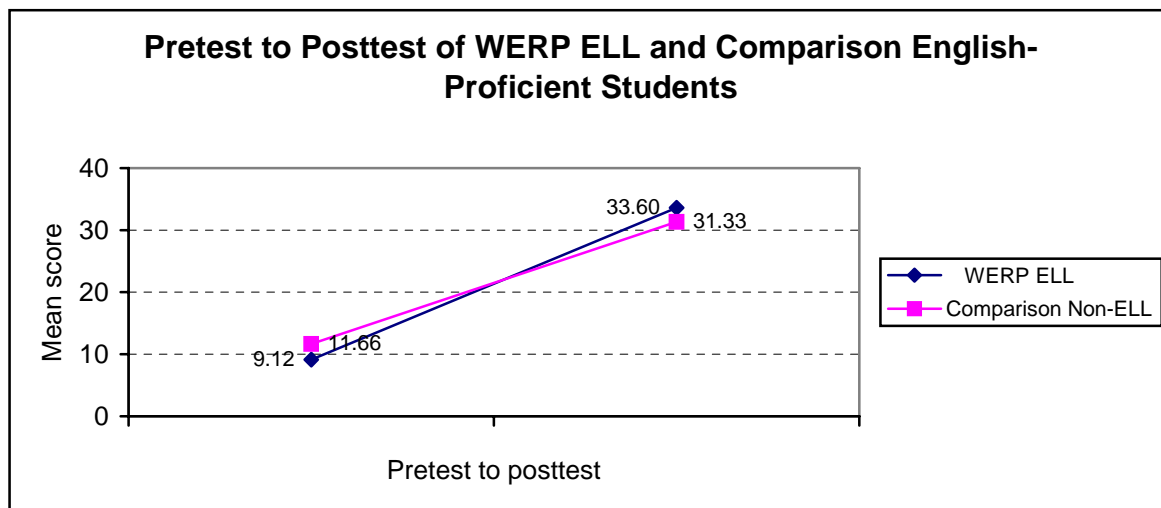


Table 22. ANCOVA of WERP ELL and Comparison Non-ELL Students

Group	N	Pretest		AdjPosttest		Adj Gain	F	p
		M	SD	M	SD			
DIBELS: Total Reading								
WERP ELL	164	9.12	6.65	33.60	12.95	24.48	8.62	.003
Comp. English-Prof.	882	11.66	8.66	31.33	12.97	<u>19.67</u>		
WERP vs Comparison						4.81***		

Note. WERP students selected with 1100 minutes (6 months) or more usage of WERP Reading Program.

* $p < .05$, ** $p < .01$, *** $p < .001$ from independent t tests comparing gains.

Figure 15. ANCOVA of WERP ELL and Comparison Non-ELL Students

WERP Usage Effects

Correlations of Program Usage and Achievement

All WERP students were included in this analysis regardless of their level of usage of the program. The total usage in minutes of the WERP software was significantly and positively correlated with posttest reading measures (Initial Sound Fluency, Letter Naming Fluency, Word Use Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, Total Reading score, and CCSA reading). The median of the significant correlations was .20. This meant that students who spent more time using the WERP software tended to have higher posttest reading achievement.

There was also a correlation between the total usage in minutes of the WERP software and the gains of students in Initial Sounds Fluency, Letter Naming Fluency, Nonsense Word Fluency, Total Reading, and CCSA reading. The median of these significant correlations was .20. These correlations suggest that those who spent more time using the WERP materials and covered more content tended have greater gains in these reading areas.

Table 23. Correlations of WERP Usage, Achievement and Gains

Measures	Total	Usage		
		Level 1	Level 2	PA
Initial Sounds Fluency	.24**	.15**	.23**	.18**
Letter Naming Fluency	.16**	.03	.40**	.12**
Word Use Fluency	.07*	.06	.02	.01
Phoneme Segmentation Fluency	.19**	.11**	.24**	.19**
Nonsense Word Fluency	.33**	.17**	.50**	.26**
Total Reading	.27**	.15**	.36**	.19**
CCSA Reading	.12**	.15**	.02	.10*
Gains Initial Sounds Fluency	.20**	-.02	.18*	.05
Gains Letter Naming Fluency	.25**	-.01	.23**	.18**
Gains Word Use Fluency	.04	.03	.09	-.13
Gains Phoneme Segmentation Fluency	.08	.13*	-.08	.12*
Gains Nonsense Word Fluency	.24**	.03	.32**	.15**
Gains Total Reading	.11*	-.01	.06	.00
Gains CCSA Reading	.22**	.07	.11	.17**

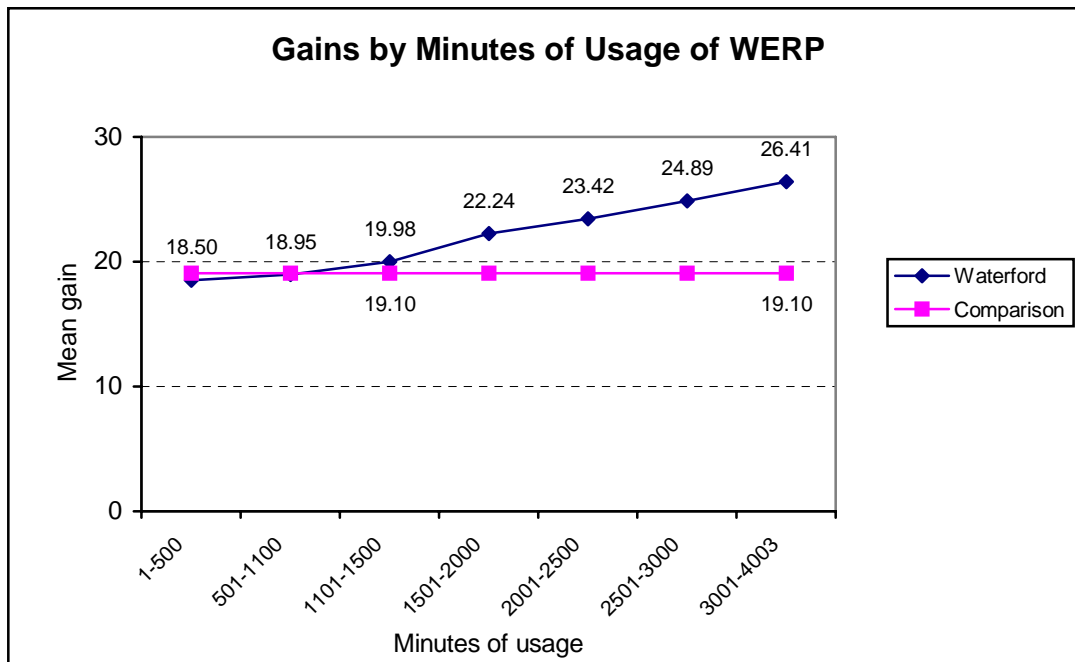
Note. PA = Phonological Awareness. Usage Level 1 = Reading Level 1 total minutes in the course; Usage Level 2 = Reading Level 2 total minutes in the course; Total = the total usage minutes of Level 1, Level 2, and Phonological Awareness. Only WERP students had usage minutes. * $p < .05$, ** $p < .01$, *** $p < .001$.

The reading achievement of the WERP students was grouped by seven levels of usage. WERP student gains at each level were compared with the comparison group gains (19.10). WERP gains increased relative to the comparison group beginning with 1 – 500 minutes of usage with significant gains starting with 1500 minutes of usage. WERP student gains over the comparison group increased by 3.14, 4.32, 5.79, and 7.31 as the usage of WERP increased. This suggests that the greater the usage of WERP content the more significant gains are made.

Table 24. DIBELS Total Reading Score by WERP Usage Level

Usage Groups	N	Pretest		Posttest		Gains
		M	SD	M	SD	
WERP 1-500 mins	153	9.72	7.14	28.22	12.86	18.50
Comparison	1211	10.22	8.29	29.32	13.33	<u>19.10</u>
WERP vs Comparison						-.60
WERP 501 –1100 mins	139	8.94	8.28	27.89	11.23	18.95
Comparison	1211	10.22	8.29	29.32	13.33	<u>19.10</u>
WERP vs Comparison						-.15
WERP 1101 – 1500 mins	76	8.67	5.99	28.65	11.46	19.98
Comparison	1211	10.22	8.29	29.32	13.33	<u>19.10</u>
WERP vs Comparison						.88
WERP 1501 – 2000 mins	84	10.05	6.40	32.29	12.19	22.24
Comparison	1211	10.22	8.29	29.32	13.33	<u>19.10</u>
WERP vs Comparison						3.14**
WERP 2001 – 2500 mins	56	11.76	7.94	35.18	12.95	23.42
Comparison	1211	10.22	8.29	29.32	13.33	<u>19.10</u>
WERP vs Comparison						4.32**
WERP 2501 – 3000 mins	94	11.35	6.96	36.24	11.58	24.89
Comparison	1211	10.22	8.29	29.32	13.33	<u>19.10</u>
WERP vs Comparison						5.79***
WERP 3001 – 4003 mins	24	13.27	8.16	39.68	12.92	26.41
Comparison	1211	10.22	8.29	29.32	13.33	<u>19.10</u>
WERP vs Comparison						7.31***

Note. * $p < .05$, ** $p < .01$, *** $p < .001$ from independent t tests comparing gains. All pretest – posttest comparisons with paired-samples t test were highly significant $p < .001$. Spearman's correlation between usage and reading gains was $r_s = .93$, $p < .001$ indicating the more usage the greater the reading gains.

Figure 16. Gains on DIBELS Total Reading Score by WERP Usage Level

SUMMARY AND DISCUSSION

A summary of the major findings of this evaluation follows. A full presentation of the findings has been presented in tables and charts of this report. The findings are unusual for the consistency across ethnic, language, gender, and achievement groups with which the gains favor the WERP.

The present study contributes to the research on the effectiveness of the WERP by disaggregating the results by gender, language and ethnicity subgroups and by considering the effects of varying dosages (minutes of usage) of the WERP.

Significant Findings

- The WERP kindergartners consistently outperformed the comparison group kindergartners on all reading outcome measures. Comparison school kindergartners did make substantial and in some cases outstanding gains from pretest to posttest. However, when WERP kindergartners were compared with comparison kindergartners, the WERP gains were substantially and significantly greater.
- Effect sizes of reading pretest to posttest gains favored the WERP kindergartners over the comparison kindergartners on all measures.
- WERP reading gains were greater for males in the WERP program than for males in the comparison group, and for females in the WERP than for females in the comparison group.
- WERP reading gains were greater for Whites, Hispanics, African Americans, Native Americans, and Asians than for their counterparts in the comparison group.
- WERP reading gains of White, African American, Hispanic, and Asian kindergartners were greater than the reading gains of White kindergartners in the comparison group.
- WERP reading gains of kindergartners with a primary home language of English, Spanish, and other languages were greater than the reading gains of their counterparts in the comparison group.
- WERP reading gains of kindergartners with a primary home language of Spanish were greater than the reading gains of English primary home language kindergartners in the comparison group. That is, WERP Spanish primary home language students who were learning English reading skills surpassed the gains of the comparison group English primary home language students.

- WERP reading gains of kindergartners in four different quartile levels of reading achievement were greater than the reading gains of comparison students with the largest gains in the top (fourth) quartile.
- WERP English language learners (ELL) reading gains were greater than the reading gains of the comparison ELL group.
- WERP ELL student reading gains were greater than the reading gains of the non-English language learners (English-proficient students) in the comparison group.
- Usage of the WERP software was found to be significantly correlated with the reading outcome measures and pretest to posttest gains in the outcome measures. This suggests that the more the student experiences the WERP content, the greater the reading gains.
- Findings from the average reading score gains by minutes of usage analyses indicate the WERP group quickly closed the gap with the comparison group and significantly outperformed the comparison group starting with 1501 minutes of usage.

APPENDIX A: READING SCORES BY SCHOOL

Table 25. DIBELS Total Reading Scores by School

WERP Schools	N	Pretest		Posttest		Gain
		M	SD	M	SD	
School A	13	10.02	5.29	33.71	8.27	23.69
School B	43	8.13	6.15	29.41	14.64	21.29
School C	13	10.03	4.72	30.51	7.50	20.48
School D	41	8.22	6.64	30.69	12.72	22.47
School E	1	.00	.00	14.00	.00	14.00
School F	3	16.87	2.14	33.87	4.70	17.00
School G	6	1.29	.71	10.07	8.02	8.78
School H	52	12.86	6.32	32.98	8.96	20.12
School J	83	11.24	6.82	35.76	11.35	24.52
School K	79	11.89	7.76	37.99	12.68	26.10
Total	334	10.62	6.98	33.59	12.42	22.97

Comparison Schools	N	Pretest		Posttest		Gain
		M	SD	M	SD	
School M	92	13.06	9.34	29.56	12.67	16.50
School N	90	12.27	8.45	34.33	13.40	22.06
School O	63	12.08	8.33	32.86	10.86	20.78
School P	95	8.10	8.00	28.65	11.80	20.55
School Q	47	10.51	6.26	38.30	11.94	27.79
School R	58	10.18	7.67	30.74	12.87	20.57
School S	102	8.14	6.74	22.56	12.03	14.42
School T	112	13.30	8.90	32.54	13.51	19.24
School U	77	7.14	6.49	19.67	12.64	12.53
School V	97	11.90	9.63	31.97	12.45	20.07
School W	49	8.51	7.44	30.74	11.85	22.34
School X	69	11.33	8.84	32.63	15.65	21.31
School Y	114	9.89	8.07	29.20	12.26	19.31
School Z	51	11.18	7.52	31.51	12.43	20.33
School AA	95	5.70	5.99	21.83	10.79	16.12
Total	1211	10.22	8.30	29.32	13.33	19.10

Note. WERP students selected with 1100 minutes (6 months) or more usage of WERP Reading Program.

APPENDIX B: READING PERCENTILES BY SCHOOL

Table 26. DIBELS Total Reading Local Percentiles by School

WERP Schools	N	Pretest		Posttest		Gain
		M	SD	M	SD	
School A	13	49.94	11.90	55.06	16.05	5.12
School B	43	41.43	17.37	47.07	23.08	5.63
School C	13	48.25	12.88	47.63	13.09	-.62
School D	41	41.07	14.93	49.78	21.14	8.70
School E	1	15.80	.00	17.60	.00	1.80
School F	3	62.13	1.45	60.60	11.68	-1.53
School G	6	20.50	4.41	17.80	14.99	-2.70
School H	52	54.58	14.88	54.38	15.84	-.20
School J	83	46.67	16.39	62.45	16.79	15.78
School K	79	49.10	16.50	63.11	19.03	14.01
Total	334	46.88	16.63	56.00	20.12	9.12

Comparison Schools	N	Pretest		Posttest		Gain
		M	SD	M	SD	
School M	92	54.85	19.87	48.66	21.45	-6.19
School N	90	51.50	18.61	58.16	22.15	6.66
School O	63	51.82	19.77	56.80	18.14	4.97
School P	95	42.79	17.23	44.66	19.89	1.87
School Q	47	46.79	15.61	61.13	15.88	14.34
School R	58	47.35	18.48	49.49	19.53	2.13
School S	102	42.76	16.17	36.76	20.08	-6.01
School T	112	52.28	17.74	52.47	19.56	.19
School U	77	40.83	17.58	32.42	20.40	-8.41
School V	97	44.58	19.99	54.28	21.53	9.70
School W	49	44.48	19.81	49.74	19.55	5.26
School X	68	45.29	22.02	49.48	22.93	4.19
School Y	114	46.92	18.16	48.33	20.59	1.42
School Z	51	48.33	16.39	51.27	21.24	2.95
School AA	95	36.75	16.30	36.62	17.91	-.13
Total	1210	46.45	18.84	47.99	21.60	1.54

Note. WERP students selected with 1100 minutes (6 months) or more usage of Waterford Early Reading Program. The pretest-posttest effect size is the mean gain divided by the standard deviation (Walberg 2001).

* $p < .05$, ** $p < .01$, *** $p < .001$ from independent t tests comparing gains.

APPENDIX C: READING SCORES BY SCHOOL FOR ALL STUDENTS

Table 27. All Students on DIBELS Total Reading Score

WERP Schools	N	Pretest		Posttest		Gain
		M	SD	M	SD	
School A	40	8.25	5.85	29.29	10.47	21.04
School B	68	7.56	6.08	27.17	13.14	19.61
School C	37	7.52	4.76	28.57	8.20	21.05
School D	49	9.00	6.89	31.81	13.00	22.81
School E	60	11.19	7.04	32.28	10.79	21.09
School F	35	9.03	6.87	26.18	11.06	17.15
School G	44	8.58	8.97	25.34	14.18	16.76
School H	66	14.88	8.83	35.34	10.14	20.46
School I	18	8.72	6.44	25.00	10.31	16.28
School J	86	11.25	6.83	35.55	11.33	24.30
School K	80	11.94	7.72	38.26	12.84	26.32
School L	53	7.91	5.60	24.19	12.42	16.28
Total	636	10.08	7.34	31.11	12.58	21.03

Comparison Schools	N	Pretest		Posttest		Gain
		M	SD	M	SD	
School M	92	13.06	9.34	29.56	12.67	16.50
School N	90	12.27	8.45	34.33	13.40	22.06
School O	63	12.08	8.33	32.86	10.86	20.78
School P	95	8.10	8.00	28.65	11.80	20.55
School Q	47	10.51	6.26	38.30	11.94	27.79
School R	58	10.18	7.67	30.74	12.87	20.56
School S	102	8.14	6.74	22.56	12.03	14.42
School T	112	13.30	8.90	32.54	13.51	19.24
School U	77	7.14	6.49	19.67	12.64	12.53
School V	97	11.90	9.63	31.97	12.45	20.07
School W	49	8.51	7.44	30.74	11.85	22.23
School X	69	11.33	8.84	32.63	15.65	21.30
School Y	114	9.89	8.07	29.20	12.26	19.31
School Z	51	11.18	7.52	31.51	12.43	20.33
School AA	95	5.70	5.99	21.83	10.79	16.13
Total	1211	10.22	8.30	29.32	13.33	19.10

Note. All WERP and comparison students included without any selection.

Table 28. Schools Ranked by Pretest Means on the DIBELS Total Reading Score

Group	School	Pretest	Posttest	Gain
WERP	School E	0.00	14.00	14.00
WERP	School G	1.29	10.07	8.78
Comparison	School AA	5.70	21.83	16.12
Comparison	School U	7.14	19.67	12.53
Comparison	School P	8.10	28.65	20.55
WERP	School B	8.13	29.41	21.29
Comparison	School S	8.14	22.56	14.42
WERP	School D	8.22	30.69	22.47
Comparison	School W	8.51	30.74	22.34
Comparison	School Y	9.89	29.20	19.31
WERP	School A	10.02	33.71	23.69
WERP	School C	10.03	30.51	20.48
Comparison	School R	10.18	30.74	20.57
Comparison	School Q	10.51	38.30	27.79
Comparison	School Z	11.18	31.51	20.33
WERP	School J	11.24	35.76	24.52
Comparison	School X	11.33	32.63	21.31
WERP	School K	11.89	37.99	26.10
Comparison	School V	11.90	31.97	20.07
Comparison	School O	12.08	32.86	20.78
Comparison	School N	12.27	34.33	22.06
WERP	School H	12.86	32.98	20.12
Comparison	School M	13.06	29.56	16.50
Comparison	School T	13.30	32.54	19.24
WERP:	School F	16.87	33.87	17.00

Note. WERP students selected with 1100 minutes (6 months) or more usage of WERP Reading Program.

REFERENCES

- Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Ellenberg, J., H. (1996). Intent-to-treat analysis versus as-treated analysis. *Drug Information Journal*, 30, 535-544.
- Finn, C. E., Rotherham, A. J., & Hokanson, C. R., Jr. (Eds.) (2001). *Rethinking special education for a new century*. Retrieved December 10, 2006, from <http://www.edexcellence.net/library/special.ed/special.ed.ch12.pdf>.
- Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills: Administration and scoring guide*. (6th ed.). Eugene: University of Oregon, Institute for the Development of Educational Achievement.
- Kirk, R. E. (1968). *Experimental design procedures for the behavioral sciences*. Belmont, CA: Brooks/Cole.
- Linn, R. L. (1981). Measuring pretest-posttest performance changes. In R. A. Berk (Ed.), *Education evaluation methodology: The state of the art*. Baltimore: The Johns Hopkins University.
- Little, R. J.A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. (2nd ed). Hoboken, NJ: Wiley.
- McKnight, P E., McKnight, K.M., & Figueredo, A.J. (2007). *Missing data: A gentle introduction*. New York: Guilford Press.
- National Association for the Education of Young Children. (1996, April). *Technology and young children – Ages 3 through 8*. [Position statement]. Retrieved December 5, 2006, from http://www.naeyc.org/resources/position_statements/pstech98.htm
- National Association for the Education of Young Children & International Reading Association. (1998, May). *Learning to read and write: Developmentally appropriate practice for young children*. [Joint position statement]. Retrieved December 10, 2006, from http://www.naeyc.org/resources/position_statements/psread2.htm.
- TUSDStats: *Useful things to know about interpreting AIMS/CCSA data*. (n.d.) Tucson Unified School District. Retrieved July 17, 2006, from <http://www.tusd.k12.az.us>

Walberg, H. J. (2001) *Final Evaluation of the Reading Initiative*. Albertson Foundation Report. Retrieved July 20, 2006, from www.waterford.org.

Waterford Institute. (2002a). Waterford Early Reading Program and No Child Left Behind. *Waterford teacher manual*. Scottsdale, AZ: Author.

Waterford Institute. (2002b). Waterford Early Reading Program and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). *Waterford teacher manual*. Scottsdale, AZ: Author.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed) New York: McGraw-Hill.

ABOUT THE AUTHORS

Stephen Powers, Ph.D., was a public school teacher for 12 years, a program evaluator with the Tucson Unified School District for 18 years and president of Creative Research Associates for 12 years. He was president of the Arizona Evaluation Network and the Arizona Educational Research Organization. As an adjunct professor in the Department of Educational Psychology of the University of Arizona and Northern Arizona University he taught research methods and statistics for over 20 years. He has published over 80 articles in research journals on topics of testing, minority assessment, personality, and student achievement.

Connie Price-Johnson, M.A., has been a research evaluator for four years with extensive experience in the evaluation of educational programs. She has expertise in second language learning and social marketing, and recently completed a national study of school districts with exemplary communication programs.