

A Study of the Correlation Between Test Gains and
Time Spent Using the Waterford Early Reading Program

Dallas (TX) I.S.D

Testing and Evaluation Division
Waterford Institute

December 1998

When a new program is implemented, the quality of implementation varies from teacher to teacher. Some teachers embrace and accept the change, while others balk at unfamiliar procedures and pedagogy. Given these circumstances, an effective program may be dismissed as ineffectual if the implementation is poor. To circumvent this problem the quality of implementation should be considered when evaluating any new program.

The Waterford Early Reading Program (WERP) is a computer-based program designed to help students acquire pre-reading and reading skills. For a complete description of WERP please refer to the background research at the beginning of this compendium.

The Waterford Early Reading Program (WERP) was implemented in three schools in Dallas, TX. This paper represents part of the program evaluation at these three sites. Other parts of the evaluation have shown that students using WERP significantly outperformed control students both on overall score and in each skill tested, except rhyme generation (see Summer/Fall 1998 Research Compendium). In addition, the Dallas Independent School District reported a significant correlation between the estimated time using WERP and word analysis gains on the ITBS (see *Evaluation of the 1997-98 Waterford Early Reading Program REIS98-326-2*, Department of Accountability and Information Systems, Dallas Public Schools). The purpose of this paper is to further explore the quality of implementation in each class at these three schools and to determine what, if any, correlation exists between the quality of implementation and pre-test to post-test gains made by each class and by each individual student.

METHODS

Sample: The sample consisted of 135 students and 12 classes.

Measurement: Quality of implementation was measured by comparing the amount of time spent using WERP (actual usage) with the recommended usage time.

Actual Usage: the computer stored Usage data as the students used the program and then downloaded onto a floppy disk at the end of the month. Specifically, these data showed the amount of time each student had used the program and the percentage of the program each student had completed. A monthly usage report with this information was also sent to each teacher. At the end of the year, the total time spent using WERP and the total percentage of WERP completed were calculated. Table 1 shows the average time spent using WERP and the average percentage of WERP completed for each class. Due to technical and scheduling problems, final data were unavailable for a few classes. Their final data were estimated by extrapolating from the previous month's data. As would be expected, usage (time spent using WERP) and progress through the program were highly correlated ($r = .91$) and varied from class to class.

Table 1: Average total time using WERP and average percentage of WERP completed

School	Average Total Time	Average % Complete
SCHOOL 1		
Class 1	1835	88%
Class 2	2033	90%
Class 3	2367	94%
Class 4	2875	100%
Class 5	2547	100%
SCHOOL 2		
Class 1	1383	58%
Class 2	1900*	85%
Class 3	1868*	88%*
Class 4	1740	88%
SCHOOL 3		
Class 1	1345	62%
Class 2	1657	79%
Class 3	829	29%

*Estimated—complete usage data unavailable

Recommended Usage: Recommended total usage for WERP is 2700 minutes. This corresponds to a session time of 15 minutes per day over a 180-day school year. This amount of time is usually sufficient for every student to complete 100% of the program. Since final data for this study was collected before the end of school, the actual recommended time was 2415 minutes.

Pre-test to Post-test Gain: Pre-test to post-test gains reflect student scores on the Waterford-Adams-Gough Reading Inventory. This test measures grade-appropriate pre-reading skills including the ability to write one's name, letter recognition, phonemic

awareness, print concepts, and decoding. Scores given are based on the total percentage of items answered correctly. Table 2 shows average pre-test scores, average post-test scores, and the average pre-test to post-test gain for each class.

Table 2: Average pre-test to post test gain by class.

School	Pre-Test % Correct	Post-Test % Correct	% Points Gain
SCHOOL 1			
Class 1	14.28%	52.10%	37.82
Class 2	37.25%	79.88%	42.63
Class 3	41.77%	79.13%	37.36
Class 4	35.15%	78.15%	43.00
Class 5	39.73%	88.58%	48.85
SCHOOL 2			
Class 1	41.28%	69.95%	28.37
Class 2	33.21%	69.60%	36.39
Class 3	33.03%	73.35%	40.32
Class 4	30.68%	67.39%	36.71
SCHOOL 3			
Class 1	10.07%	39.94%	28.77
Class 2	10.95%	52.59%	41.64
Class 3	22.25%	51.02%	29.87

RESULTS

Pearson correlation coefficients were calculated to determine what, if any, correlation existed between time spent using WERP and pre-test to post-test gain. Table 3 shows the coefficients obtained.

Table 3: Correlation coefficients using class averages and using individual student scores

	Correlation coefficient (r)
Class Usage to Gain	0.79**
Student Usage to Gain	0.15*

*Correlation is significant at the .05 level (2-tailed) **Correlation is significant at the .01 level (2-tailed)

The data indicate a statistically significant high correlation ($r = .79$) between the total time spent using WERP by a class and the average pre-test to post-test gain. A statistically significant correlation ($r = .15$) also exists between the time each student spent using WERP and his or her pre-test to post-test gain. The lower student correlation can likely be attributed to increased variance among students, but may also be affected by a ceiling effect on the test. That is, if a student receives a score of 95% on the pre-test, the maximum possible gain is 5 points (post-test score of 100%) regardless of how much he or she uses the program. If such a student were to use the program as recommended, it would reduce the magnitude of the usage to gain correlation. The correlation for classes is much less likely to be affected by this phenomenon because the highest pre-test average for a class was 41.77%.

The trend indicated by the Pearson coefficients can easily be seen when the classes and students are placed into usage groups (with a new group every 500 minutes) and the average gain for each group is plotted. The gains made by each of these groups were also compared to gains made by a control group, that is, students and classes not using WERP (Figures 2 and 3).

Figure 2: Class usage vs. gain by category

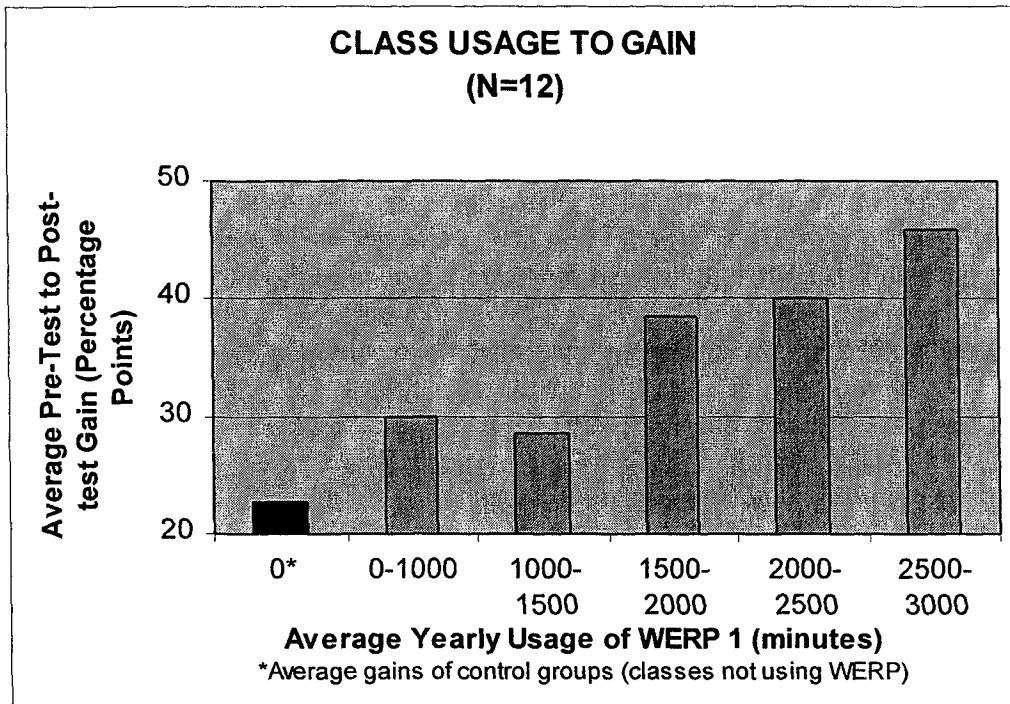
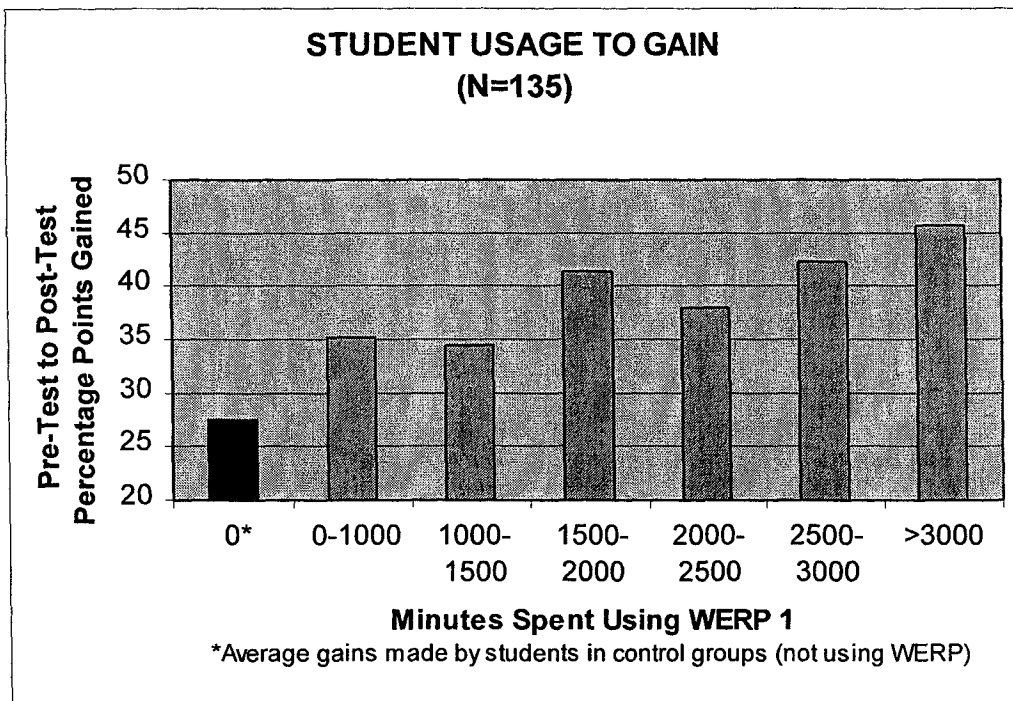


Figure 3: Student usage vs. gain by category



DISCUSSION

As with any new program, the quality of implementation of the Waterford Early Reading Program in these three Dallas sites varied from class to class. The data indicate that increased usage of WERP corresponded to greater test gains. This was true for both classes and individual students. While these data do not necessarily present a causal relationship, the results are encouraging. Regardless of the cause, classes and students that used WERP as recommended experienced more pre-test to post-test gain than those who did not. These results should also engender vigilance on the part of administrators and evaluators. Since higher usage corresponds to better test gains, maintaining high usage should be a priority for all sites using WERP. Furthermore, since low usage corresponds to low test gains, any evaluation of the Waterford program should include an assessment of the time spent using the program. Finally, continued investigation is recommended to determine if the correlation of usage to gain found in this study is consistent for all sites using WERP.