WILSON LITERACY SOLUTIONS

EVIDENCE OF EFFECTIVENESS

Wake Forest University, 2002: **Data Analysis**

Pre-intervention and post-intervention test data was analyzed with the assistance of Dr. Frank Wood et al at Wake Forest University School of Medicine.

Results indicate there was ***improvement across all Woodcock Reading Mastery subtests***. Further analysis was conducted on difference by gender, IQ, severity, grade, and school district.

Data Analysis of the Wilson Reading System®

Wilson Language Training collects pre- and post-test data to document the efficacy of the Wilson Reading System®.

The data have been analyzed with the assistance of Frank Wood, Ph.D. et al of Wake Forest University School of Medicine, Winston-Salem, NC.

2002

Data Analysis

The data presented below were collected from Wilson tutors in training at multiple sites across the US. Prospective tutors are required to undergo supervision while providing Wilson Reading System® services to a child deemed by the respective school system to need remedial assistance. Pre-intervention and post-intervention test data are collected by Wilson to document efficacy of the Wilson program. The data have been analyzed with the assistance of Frank Wood et al. of Wake Forest University School of Medicine.

Test data usually consist of Woodcock-Johnson Tests of Achievement - Revised (or, more recently, 3rd Edition) or Woodcock Reading Mastery test scores from the reading subtests of those measures. Some school districts, however, use other measures for the pre- and post-intervention assessment (e.g., Wechsler Individual Achievement Test).

The data reported herein were compiled from two cohorts, one having received intervention during the 1999-2000 school year, the second during the 2000-2001 school year. Only those students for whom the Woodcock Reading Mastery test (WRMT) was administered were included in the analyses.

Cohort Comparison.

Mean Woodcock Reading Mastery scores for each cohort as well as the group as a whole (with the two cohorts combined) are presented in Table 1. Scores are age-referenced standard scores. There were 405 students with pre-test scores available for analysis in the combined cohort, with 208 students in cohort 1 and 197 students in cohort 2. The number of students who received post-testing was, in most cases, lower (although not always as some students had post-test scores available yet did not have pre-test scores available). There were 176 students from cohort 1 and 198 students from cohort 2 with post-test scores available and a total of 374 students with post-test scores for the combined cohorts.

Table 1. Mean WRMT scores by cohort.

Subtest	**Combined Cohorts**	**Cohort 1**	**Cohort 2**
Word Identification			
Pre-test	83.71	83.00	84.48
Post-test	88.99	89.67	88.37
Word Attack			
Pre-test	87.73	87.74	87.71
Post-test	98.85	98.98	98.73
Passage Comprehension			
Pre-test	86.21	86.45	85.95
Post-test	92.49	92.86	92.17
Basic Skills Cluster			
Pre-test	84.42	83.93	84.94
Post-test	92.39	92.93	91.90
Total Reading Cluster			
Pre-test	83.83	83.38	84.31
Post-test	89.61	90.03	89.25

The mean pre-intervention and post-intervention age for both cohorts was 10 and 11 years, respectively. Full-scale IQ scores were available for 133 of the cohort 1 subjects and for 44 of the cohort 2 subjects. Mean IQ scores were 97.7 and 96.4, respectively. T-tests were computed comparing cohort 1 WRMT mean scores and cohort 2 WRMT mean scores with no statistically significant differences noted (all p's >.18).

Change scores for each subject were computed for each of the WRMT subtest and cluster scores (post-test score - pre-test score) and mean change scores generated. T-tests comparing each cohort's change scores also failed to reveal significant differences, suggesting that the two cohorts were quite similar in pre-test and post-test scores as well as changes in those scores (all p's >.17). Thus, it was deemed feasible to combine the two cohorts for further analyses.

Gender Comparison.
Possible differences between male and female subjects was explored by a similar procedure whereby subtest score means were computed for each of the WRMT subtest and cluster scores and t-tests were employed to determine whether or not the genders differed on any measure. (Note that information regarding gender was missing for three individuals and those subjects were excluded from the analysis.) Of note, full-scale IQ scores were available for 97 males and for 80 females (means of 98.27 and 96.23, respectively). WRMT means are presented in Table 2. There were 219 males and 183 females with pre-test scores while 209 males and 162 females had post-test scores.

Table 2. Mean WRMT scores by gender.

Subtest	**Entire Sample**	**Males**	**Females**
Word Identification			
Pre-test	83.71	83.81	83.58
Post-test	88.99	88.30	89.90
Word Attack			
Pre-test	87.73	88.37	86.95
Post-test	98.85	99.23	98.37
Passage Comprehension			
Pre-test	86.21	85.95	86.55
Post-test	92.49	92.42	92.78
Basic Skills Cluster			
Pre-test	84.42	84.72	84.05
Post-test	92.39	92.09	92.80
Total Reading Cluster			
Pre-test	83.83	83.87	83.79
Post-test	89.61	89.25	90.16

As in the earlier analysis, none of the t-tests yielded statistically significant results (all p's >.11). Again, change scores were computed (post-test score - pre-test score) for each subject and mean change scores for each sex compared for significant differences. No statistically significant differences emerged (all p's >.32) and therefore it was deemed appropriate to combine the groups for further analyses.

Efficacy of the Wilson Reading System® (WRS).

A glance at the entire sample's means in both Tables 1 and 2 reveals a notable improvement in WRMT scores from pre-testing to post-testing, especially for the Word Attack subtest. As predicted, the following subtests and cluster scores improved to a statistically significant degree: 1) Word Identification, $t(350) = -12.35$, $p<.0001$; 2) Word Attack, $t(348) = -22.56$, $p<.0001$; 3) Passage Comprehension, $t(348) = -13.72$, $p<.0001$; 4) Basic Skills Cluster, $t(358) = -19.70$, $p<.0001$; 5) Total Reading Cluster, $t(348) = -15.69$; $p<.0001$.

These results indicate that there was indeed improvement across all WRMT subtests, more than would be expected from chance variation. Indeed, the Word Attack subtest alone yielded an 11-point average improvement in age-referenced standard scores.

Effects of IQ on student gains.

Most students received the Wechsler Intelligence Scale for Children - 3rd Edition (WISC-III) although other measures were sometimes used as well (e.g., the Otis-Lennon, Differential Ability Scales, etc.). In cases where a global, or full-scale score was reported, it was this score that was used for data analysis, regardless of the measure used for testing.

To examine any group differences on the basis of IQ, the sample was divided into six subgroups based on IQ score as follows: Group 1 - IQ scores between 70 and 79 (n=12); group 2 - IQ scores between 80 and 89 (n=36); group 3 - IQ scores between 90 and 99 (n=44); group 4 - IQ scores between 100 and 109 (n=38); group 5 - IQ scores between 110 and 119 (n=18); group 6 - IQ scores 120 or above (n=10). Means for each of the six groups are reported in Table 3.

Table 3. Means by IQ.

Subtest	**Group 1**	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**
Word Identification						
Pre-test	67.00	79.84	81.07	83.11	88.28	92.6
Post-test	76.33	85.33	88.45	89.03	92.73	96.38
Word Attack						
Pre-test	76.15	87.08	85.93	89.08	90.06	91.6
Post-test	92.58	97.97	98.18	99.06	101.27	102.31
Passage Comprehension						
Pre-test	70.46	81.76	85.09	88.21	90.78	94.7
Post-test	77.33	84.97	92.5	94.3	99.2	105.0
Basic Skills Cluster						
Pre-test	70.75	81.62	82.14	84.37	87.83	91.5
Post-test	81.75	89.63	91.76	92.3	95.6	99.0
Total Reading Cluster						
Pre-test	67.83	79.81	81.49	84.29	88.22	92.7
Post-test	75.42	84.72	87.84	90.24	94.87	99.46

Mean IQ scores for each of the six groups were significantly different from each other, $F(5, 170) = 661.77$, $p<.0001$ (all follow-up Tukey t-tests were statistically significant). Each group's improvement was then examined for all five WRMT measures. All six groups saw statistically significant gain in their Basic Skills and Total Reading clusters from pre-test to post-test. Similarly, all six groups had statistically significant gains in Word Attack from pre-test to post-test. The Word Identification improvement from pre-test to post-test was significant for all

groups with the exception of group 1 where the difference approached significance, t(8) = -2.15, p = .06. Similarly, there was only one group (group 2) for whom the improvement in passage comprehension scores was not significantly different, t(27) = -1.72, p=.097.

While a cursory inspection of the means from pre-testing to post-testing suggests a tendency for the lower IQ groups to make more gains in Word Attack than the higher IQ groups (e.g., the mean change in Word Attack score for group 1 is 17 standard score points while the mean change in Word Attack for group 6 is 12 standard score points), these differences in change scores did not reach statistical significance. These data nonetheless refute the notion that children with lower IQ scores might not benefit from remediation as well as do higher IQ children. Indeed, for children with IQ scores between 70 and 79, the average Word Attack score fell in the borderline range prior to WRS intervention while the average post-test Word Attack score reached the fully average range. Gains in Word Identification were also noteworthy, with the average pre-test score falling in the impaired range (67) while the average post-test score reached the borderline range (76).

Effects of severity on student gains.

To examine group difference in terms of severity, subjects were divided into three groups on the basis of the pre-test Total Reading Cluster as follows: if cluster scores were 79 or below, students were placed in group 1 (n=114); if cluster scores fell between 80 and 89, students were placed in group 2 (n=170); cluster scores falling above 90 resulted in students being placed in group 3 (n=122). The maximum Total Reading Cluster (TRC) standard score was 109.

Table 4. Means by severity.

Subtest	**TRC=1**	**TRC=2**	**TRC=3**
Word Identification			
Pre-test	69.98	85.05	94.80
Post-test	80.45	89.34	98.07
Word Attack			
Pre-test	79.91	88.66	94.41
Post-test	93.51	99.05	104.56
Passage Comprehension			
Pre-test	73.65	87.07	96.84
Post-test	83.55	93.28	101.27
Basic Skills Cluster			
Pre-test	72.70	85.47	93.99
Post-test	84.90	92.47	100.68
Total Reading Cluster			
Pre-test	69.96	85.08	95.06
Post-test	80.22	90.33	99.0

As expected based on group composition, mean pre-test TRC standard scores were significantly different between the three groups. Within each level of severity, statistical analysis indicated that there was significant improvement on all of the WRMT measures (all p's <.0001).

Analysis of differences between the three groups revealed that each of the three groups significantly differed in the degree of change from pre-test to post-test for the Word Identification, F(2, 348) = 3.40, p<.04, and Passage Comprehension, F(2, 346) = 3.97, p<.02,

subtests. Thus, it is not surprising that the groups also differed in the degree of change from pre-test to post-test for the Total Reading Cluster score, $F(2, 346) = 5.32$, $p<.01$. Follow-up t-tests (Tukey's HSD) indicate statistically significant difference between groups 1 and 3 in terms of their improvement from pre-test to post-test with an average of 5.59 standard score points improvement in group 1's Word Identification performance and an average of 3.24 standard score points improvement in group 3's performance. Similarly, in terms of Passage Comprehension performance, group 1 made an average 10.76 standard score point gain while group 3 made only a 4.41 standard score point improvement. Thus, it is not surprising that the most severe group saw greater improvement (6.76 points) in their Total Reading Cluster than did the least severe group (3.94 points).

Differences by grade.

Efficacy of the Wilson Reading System® was also examined within and across grades. Data were obtained for only a handful of 1st and 2nd graders so that those grades were not analyzed due to too few subjects. However, students in grades three through seven/eight were sufficient in number to provide meaningful comparisons. Seventh and eighth grade samples were combined. Mean subtest scores are reported in Table 5.

Table 5. Means by Grade.

Subtest	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7/8**
Word Identification	*(n=52)*	*(n=106)*	*(n=88)*	*(n=54)*	*(n=71)*
Pre-test	90.51	88.55	84.74	81.22	75.83
Post-test	94.28	93.37	88.27	84.35	82.51
Word Attack					
Pre-test	92.5	90.0	87.93	85.88	83.31
Post-test	101.82	100.94	99.19	95.67	95.49
Passage Comprehension					
Pre-test	91.15	90.24	87.23	82.65	81.14
Post-test	97.12	96.71	92.64	87.15	88.81
Basic Skills Cluster					
Pre-test	90.21	87.99	85.16	82.72	77.73
Post-test	96.44	96.10	92.11	88.08	87.05
Total Reading Cluster					
Pre-test	89.81	88.37	85.0	81.17	76.72
Post-test	94.90	94.25	89.49	84.46	84.39

Comparison of pre-test and post-test scores within each grade level revealed statistically significant improvements among all five WRMT measures for each grade level (all p's $<.001$).

Further analyses revealed no significant differences between the groups in terms of degree of change from pre-test to post-test. In other words, 3rd grade students did not exhibit more or less gain in scores (than would be expected by chance) than did 8th grade students for example. An important observation, therefore, is that it appears that older students benefit from WRS intervention just as much as do younger students.

Differences by school district.

To examine the efficacy of the Wilson Reading System® in an inner-city school setting, we selected from the larger sample, those students from the following school systems for whom we

had at least the Word Identification, Word Attack, and Passage Comprehension subtests from the Woodcock Reading Mastery Test - Revised (NU/ASSIST): Boston, MA; Brooklyn, NY; Lynn, MA. Detroit, Albuquerque, and Newark schools were considered although no student from either of these systems had all three subtests or, in some cases, other data were missing (e.g., birthdates necessary to compute standard scores from raw scores). There were a total of 40 students from these school districts who met these criteria for inclusion in the analysis.

Table 6 reports the age-referenced standard score means of each subtest, as well as two clusters, for both the small sample of children in the inner-city schools and for the much larger, nationwide, sample. The change from pre-test average to post-test average is reported in parentheses.

Table 6. Means for inner-city schools.

Subtest	**Entire Sample**	**Inner-City Schools**
Word Identification		
Pre-test	83.71	79.38
Post-test	88.99 (5.28)	84.98 (5.6)
Word Attack		
Pre-test	87.73	84.68
Post-test	98.85 (11.12)	97.45 (12.78)
Passage Comprehension		
Pre-test	86.21	80.80
Post-test	92.49 (6.28)	87.60 (6.8)
Basic Skills Cluster		
Pre-test	84.42	80.45
Post-test	92.39 (7.97)	89.40 (8.95)
Total Reading Cluster		
Pre-test	83.83	79.05
Post-test	89.61 (5.78)	85.23 (6.18)

Statistical analysis revealed significant change in scores from pre-test to post-test for the inner-city school students across all five WRMT measures (all p's$<.001$).

Comparison of change scores for the smaller, inner-city group and the larger, nationwide, sample, reveals a virtual one-to-one correspondence in age-referenced standard score change for each of the different subtest and cluster measures. In other words, a five point standard score improvement in word identification is noted for both the small sub-sample of inner-city subjects and the larger, nationwide, group. For two measures (Word Attack and Basic Skills Cluster), the average improvement in age-referenced standard scores is slightly better for the inner-city students than for the larger group of students.