

Evaluation Report
SAT-9 Reading Test Achievement Levels
District of Columbia Public Schools
In2Books Classrooms Compared to Other DCPS Classrooms
2003-2004 School Year

Susan R. Goldman, Ph.D.
Co-Director
Center for Learning, Instruction, and Teacher Development
University of Illinois at Chicago

INTRODUCTION

In2Books is a literacy program designed to foster reading and writing skills in elementary school students. It is based on the premise that effective literacy instruction uses reading and writing for authentic communication purposes. Students are paired with electronic adult pen pals. Students and their pen pals read the same books and write letters to one another about the books. In2Books provides ongoing professional development for teachers implementing the program. The program has been in operation in District of Columbia Public Schools since 1997.

During the 2003-2004 academic year, an evaluation to study of the impact of the program on students in grades 2, 3 and 4 was undertaken. One aspect of this study examined performance on the district-administered reading achievement test, Stanford Achievement Test, Ninth Edition (SAT-9). The overall purpose of the analyses of the SAT-9 reading achievement test data was to compare the performance of students in the In2Books program with students in classrooms not participating in the program. This report provides the results of those analyses.

DATA ANALYSIS

Students in the district were administered the SAT-9 reading achievement test in March, 2004. Scale scores on the SAT-9 were made available to the independent evaluator in Fall, 2004. Analyses were conducted to compare performance on the SAT-9 by students in grades 2, 3 and 4 who were in classrooms taught by five types of teachers:

1. **Veteran** In2Books teachers were those teachers who had been implementing the program for two or more years.
2. **First-Year** In2Books teachers were implementing the program for the first time during the 2003-2004 school year and were in schools where there were also Veteran teachers who had implemented In2Books in previous years.
3. **Experimental** first year In2Books teachers were implementing the program for the first time during the 2003-2004 school year, were in schools that had not previously implemented In2Books in any classroom, and were randomly selected to be in the Experimental group for analyses that were part of a quasi-experimental research design to examine various aspects of program implementation effects on student achievement.
4. **Control** teachers were teachers who volunteered to implement In2Books during 2003-2004 but were randomly selected to wait a year to implement the program so that they could be part of the quasi-experimental research described in (3) above.

5. **Non-In2Books** teachers were teachers in DCPS who were not participating in the In2Books program in any way, did not implement In2Books during the 2003-2004 school year, and were not part of a control group of teachers identified for the quasi-experimental design study being conducted [see (3) and (4) above].

Table 1 provides the distribution of the number of students and teachers whose scores were present in the database of SAT-9 scores supplied by DCPS for each of these five groups.

Group 1: Veteran I2B Classrooms	N of Students
Grade 2 (11 classrooms)	163
Grade 3 (26 classrooms)	462
Grade 4 (21 classrooms)	390
Group 2: First-Year I2B Classrooms	N of Students
Grade 2 (4 classrooms)	65
Grade 3 (13 classrooms)	226
Grade 4 (13 classrooms)	210
Group 3: Experimental Classrooms	N of Students
Grade 2 (8 classrooms)	141
Grade 3 (20 classrooms)	344
Grade 4 (11 classrooms)	218
Group 4: Control Classrooms	N of Students
Grade 2 (10 classrooms)	153
Grade 3 (16 classrooms)	283
Grade 4 (15 classrooms)	269
Group 5: Non I2B Classrooms	N of Students
Grade 2*	1865
Grade 3*	3121
Grade 4*	3648

*The specific number of Non-In2Books classrooms per grade level cannot be determined from the DCPS database because data from numerous schools are labeled "No Name Given." Therefore only numbers of students are provided.

Two issues regarding the available data set should be noted. First, because of the fact that a number of classrooms in the DCPS database were labeled "No Name Given" for the teacher name, these classrooms could not be included in the data analysis because it could not be determined whether the scores of these groups of students were from In2Books or non-In2Books classrooms. In all, 527 scores (4.2%) of the student cases had to be omitted from the analyses because of this lack of clarity in the database.

Second, with respect to the Grade 2 data, it is the policy of DCPS to allow individual schools the choice of whether or not to administer the SAT-9 test to Grade 2 (and Grade 1) students. As a result of this policy, there were considerably fewer scores in the analysis for grade 2 students than for students in grades 3 and 4; only 41% of the potential grade 2 In2Books classrooms in the DCPS database actually administered the test and were therefore able to be included in the analysis. Furthermore, because of the problem with identifying non-In2Books classrooms owing to the “No Name Given” cases (described above), it was impossible to determine the exact percentage of Non-In2Books grade 2 classrooms not tested. But, given the other numerical trends in the data set, it is most likely close to the approximately 60% of In2Books classrooms not tested.

Statistical Analyses. The results were analyzed in a one-way analysis of variance within each grade with a five-level grouping variable. In each case, there was a statistically significant main effect of group. Tukey’s post hoc tests were used to determine whether differences between means were statistically significant. Experiment-wise error rate was set at $p < .05$ using the Bonferroni adjustment for multiple tests among means.

SUMMARY OF THE SAT-9 READING TEST RESULTS

At all grade levels, significant differences favoring students of I2B teachers as compared with students not in the program were found. Similarly, at all grade levels, students in In2Books control classrooms and students not in the program performed at statistically equivalent levels and below that of students in veteran schools.

Performance comparisons of the Experimental group with the other four groups varied depending on the grade level. In grade 3, students of Experimental teachers performed at levels equivalent to those of students of Veteran I2B teachers. These two groups significantly outperformed the other three groups, which were equivalent. In grades 2 and 4, students of First year implementers in veteran schools outperformed all other groups, students of Veteran I2B teachers outperformed those of Experimental teachers; and Experimental, Control and Students not in the program performed at statistically equivalent levels.

Table 2 provides a summary of the means across the five groups in each grade and the statistical support for the summarized results.

Table 2. Mean Scale Scores on Reading Test of the SAT-9

Grade Level	Veteran I2B	First Year I2B— Vet Schools	First Year I2B— Experimental	Control I2B	Non-I2B
2	584.45	602.12	570.58	578.09	578.66
3	626.91	601.96	620.13	605.78	607.65
4	637.34	648.34	627.02	623.56	626.78

Patterns found for the key comparisons at each grade level were as follows:

Grade 2: Students in First Year I2B—Veteran Schools scored significantly higher than each of the other 4 groups. Students in Veteran In2Books classrooms scored higher than Control and Non-I2B classrooms. Experimental classrooms scored below the level of Veteran I2B and were statistically equivalent to Control and Non-I2B classrooms.

Grade 3: Veteran In2Books and First Year—Experimental classrooms were statistically equivalent and each was significantly higher than each of the other three groups. There were not statistically significant differences among the First Year I2B—Veteran Schools, Control, and Non-I2B classrooms.

Grade 4: Students in First Year I2B—Veteran Schools classrooms scored significantly higher than each of the other 4 groups. Veteran In2Books classrooms scored significantly higher than Experimental, Non-In2Books and Control classrooms, which did not differ from one another.

CONCLUSIONS

Overall, analyses of the data from the DCPS administered SAT-9 Reading Test show clear benefits of the In2Books program. At each grade level—2, 3, and 4—students in one or more of the In2Books groups (Veteran I2B, First Year I2B—Veteran Schools, or First Year I2B—Experimental—scored significantly higher than the non-program groups (Control and Non-I2B), and the two non-program groups performed at statistically equivalent levels.

Specific patterns of advantage for the three groups of In2Books students were varied across grade levels, but there are indications that In2Books is especially advantageous for students in classrooms where teachers have participated for multiple years or are in a school that has participated in the program for multiple years. In addition, results at the grade 2 level have a degree of instability owing to selection issues related to district policy on taking the test.

Because these analyses indicate the benefits of the In2Books program experience for DCPS students at grades 2, 3, and 4, it would be useful for future evaluations to track individual students across multiple years to note patterns associated with multiple years of student experience in the program. Such an analysis at the student level would also enable multiple year inquiry into the patterns of achievement in the Experimental and Control group students, which are especially interesting because of the random assignment of these students. During 2003-2004, the Experimental group students were statistically higher in SAT-9 Reading achievement scores than the Control group at Grade 3 and equivalent to the Control students at grades 2 and 4.