## Comparative Effectiveness of *Scott Foresman Science*: *A Report of Randomized Experiments in Five School Districts*

We investigated whether *Scott Foresman Science* is more effective than current science programs in five diverse sites. Although we found no evidence that it improved science achievement beyond the regular programs, boys and girls performed equally well, whereas the control group boys outperformed girls. Our results also show that under some conditions the program can enhance reading achievement.

Introduction. Pearson Education contracted with Empirical Education Inc. to conduct randomized experiments to determine the effectiveness of its *Scott Foresman Science* products *(SFScience)* compared to the elementary science programs already in place in five geographically and demographically diverse sites. We compared science and in reading outcomes for classes using the *SFScience* curricular materials and control classes using each district's current materials.

Scott Foresman Science, a year-long curriculum intended for daily use, provides a sequence of structured and supportive inquiry activities and text materials to develop students' independent investigative skills. Science kits contain materials for hands-on activities, while Leveled Readers help the teacher differentiate instruction and provide reading support at, below, and above grade level. During the half-day training, teachers learned how the materials were to be used and how much was to be covered. Control teachers typically used state, district, and teacher developed materials, magazines, videos, online resources, and older science texts for science instruction.

Findings for Science. Overall, we found that students in the *SFScience* classrooms improved in science achievement at the same rate as the students in the established program. The following graph shows the comparison combining the results from all five districts. The set of bars on the left indicate the pre and post results for the control and *SFScience* groups. The bar graph on the right shows the results for control and *SFScience* as predicted by our statistical model that took pretest and other factors into account. The overlapping confidence intervals at the top of the bars indicate there is no statistical difference between the two groups.
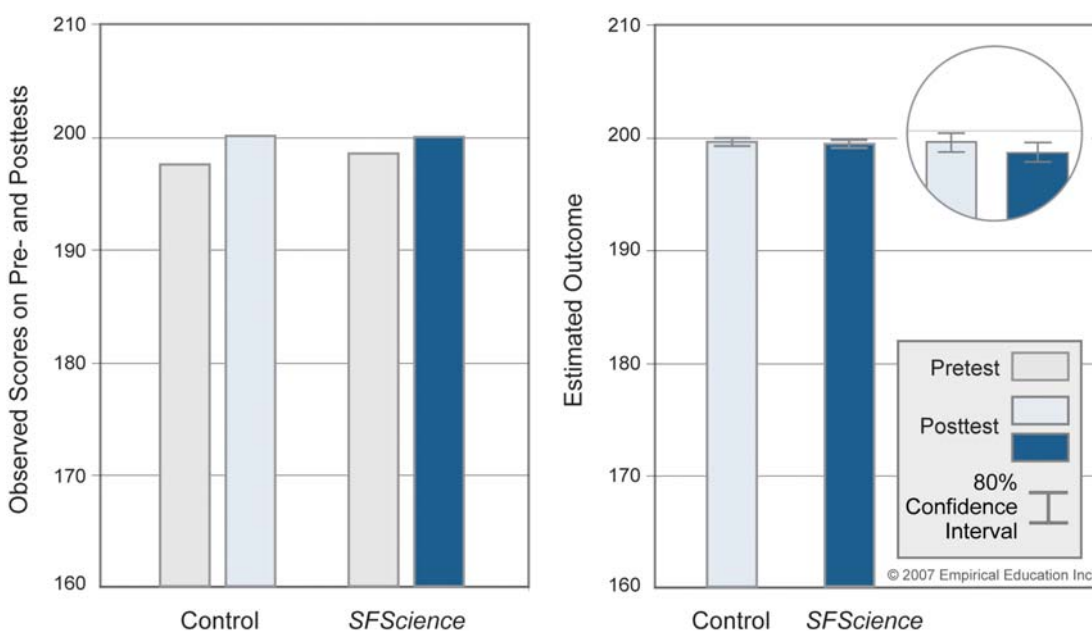
Figure 1. Impact on Science Achievement: Unadjusted Pre- and Posttest Means for Control and *SFScience* (Left); Adjusted Means for Control and *SFScience* (Right)

One interesting finding is that boys in the control group outperformed girls in science, whereas in *SFScience* boys and girls performed equally well.

Using data from observations, interviews, and surveys, we monitored the overall level of implementation at each site and we considered classroom process measures such as the amount of instructional time teachers devoted to science and the extent of inquiry teaching. These variables appeared not to impact science achievement. Nor did we find differences across grade or prior achievement levels or teacher experience.

Findings for Reading. Because *SFScience* provides a significant reading component, we also determined the amount of reading improvement that can be accounted for by the science program. Figure 2 compares the overall results for reading across the five sites and in combination. The combined results are positive, and two sites show positive results within the 80% confidence interval. The point-and-whiskers shows our estimates (the center points) within an interval representing 80% confidence; that is, if we consider each site separately, we can be 80% sure that the true value of the impact lies within the interval. In two sites, *SFScience* caused a small increase beyond expected gains for the schools' reading program by itself. When all sites are combined, however, this positive difference is insufficient to give us confidence that the difference was not due to chance.

Overall, it appeared that sites were more successful in teaching reading than science, reflecting relative emphasis on the two subjects. It is also relevant that this was the first year of use of *SFScience* and the teachers' initial unfamiliarity may have affected implementation, which differed at each individual site.
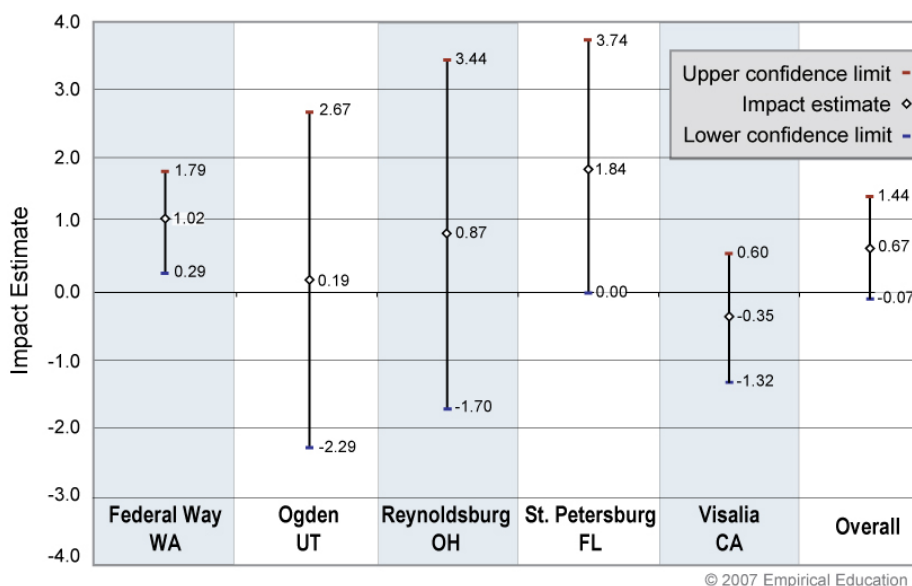
Figure 2. Estimated Reading Impacts Across Districts

© 2007 Empirical Education Ir

Our conclusion is that *SFScience* stands up to other science programs in schools. Educators may find the program attractive in the equal help it gave to boys and girls compared to other programs in place. The reading component's capacity for improving reading achievement under some conditions points to a potentially important strength of the program.

Design and Analysis. This study was a multi-site group randomized trial in which volunteer teachers within each district were assigned by coin toss to use the new program or continue with their current program for approximately one school year. Statistical analyses were based on 92 teachers/classes (46 *SFScience* and 46 control) and 2,638 students in grades 3–5. The primary outcomes, as well as pretest measures, are student-level test scores on the Northwest Evaluation Association (NWEA) test in two areas: Science Concepts and Processes and Reading. The mean impact is estimated using multi-level models. The impacts were estimated using multi-level models run in SAS PROC MIXED.

To read the complete report, contact:

Empirical Education Inc.
www.empiricaleducation.com
425 Sherman Avenue, Suite 210
Palo Alto, CA 94306
(650) 328-1734

EMPIRICAL EDUCATION RESEARCH SUMMARY