

Abstract Title Page.

Title:

Curriculum Matters: Evidence from a Randomized Control Trial of Four Elementary School Math Curricula

Author(s):

Roberto Agodini, Ph.D., Mathematica Policy Research
Barbara Harris, Mathematica Policy Research

Abstract Body

Background / Context:

National achievement data show that elementary school students in the United States, particularly those from low socioeconomic backgrounds, have weak math skills (National Center for Education Statistics 2009). In fact, data show that, even before they enter elementary school, children from disadvantaged backgrounds are behind their more advantaged peers in basic competencies such as number-line ordering and magnitude comparison (Rathburn and West 2004). Furthermore, after a year of kindergarten, disadvantaged students still have less extensive knowledge of mathematics than their more affluent peers (Denton and West 2002).

What is taught to students and how it is taught may be important factors in a school's ability to improve student math achievement; however, as Hiebert and Grouws (2007) explain, research has not identified which specific features of teaching are most effective at developing math skills. As of October 2009, the What Works Clearinghouse (WWC) had reviewed 315 studies of interventions designed to improve math achievement of elementary school students (<http://ies.ed.gov/ncee/wwc/>). Only 10 of those studies (2 of which involved using an experimental design) were judged as providing evidence that was useful for assessing the effectiveness of the interventions examined. Other reports also point to the lack of rigorous evidence on the effectiveness of various instructional approaches (National Mathematics Advisory Panel 2008; National Research Council 2004).

As Hiebert and Grouws (2007) also explain, although it would be useful to understand which features of teaching help develop student math skills, individual features typically function within a system, such as a curriculum, and the effects of each feature may depend on the system in which it functions. The potential interdependence among teaching features points to the need to study the effects of entire curricula, particularly comparing the effects of different approaches to packaging together the various teaching features.

Purpose / Objective / Research Question / Focus of Study:

This study, sponsored by the Institute of Education Sciences (IES) in the U.S. Department of Education, examines whether some early elementary school math curricula are more effective than others at improving student math achievement in disadvantaged schools. A small number of curricula, which are based on different approaches for developing student math skills, dominate elementary math instruction—7 curricula make up 91 percent of those used by K–2 educators, according to a 2008 survey (Resnick et al. 2010). However, as described above, little rigorous evidence exists to support one approach over another.

The main questions addressed by the study are:

- What are the relative effects of the study's four math curricula on math achievement of first- and second-graders in disadvantaged schools?
- Are the relative curriculum effects influenced by school and classroom characteristics?

The study is addressing these questions by examining the relative effects of four diverse curricula selected through a competitive process, recruiting schools to participate in the study,

having publishers provide schools and teachers with curriculum materials and training, assessing student math achievement, and evaluating the effects of the curricula (including materials, training, content, and pedagogy) on student achievement. Other data (including school records, teacher surveys, and classroom observations) also were collected to examine whether relative curriculum effects are influenced by context and to conduct mediational analyses.

Setting:

The 12 participating districts are geographically dispersed across 10 states and in all four of the Census Bureau–designated regions of the country. The districts also differ in terms of urban status—3 districts are in an urban area, 5 are in a suburban area, and 4 are in a rural area.

When compared to the average U.S. district, those that agreed to participate have a higher fraction of schoolwide Title I eligible schools, students eligible for free or reduced-price meals, and minority students. A similar pattern exists when comparing U.S. elementary schools with those that agreed to participate.

Population / Participants / Subjects:

The study team recruited 110 schools from the 12 districts to participate in the study. Each district contains at least four participating elementary schools so all four curricula could be implemented in each district during the 2006-07 or 2007-08 school year. During the first year of study participation, 39 schools implemented the curricula only in the first-grade, 70 implemented them in both the first and second grades, and 1 school implemented only in the second grade.

As a result, 109 schools were included in the first-grade analysis, and 71 in the second-grade analysis. In terms of classroom and student sample sizes, the first-grade analysis includes 461 classrooms and 4,716 students; the second-grade analysis includes 328 classrooms and 3,344 students. A random sample of about 10 students in each classroom was administered the study’s math assessment.¹

Intervention / Program / Practice:

Four curricula were selected for inclusion in the study using a competitive process in which developers and publishers of early elementary school math curricula were invited to submit a proposal to include their curricula in the evaluation. A panel of experts in math and math instruction reviewed the proposals and recommended to IES curricula suitable for the study.

The following four curricula were selected for the evaluation:

- ***Investigations in Number, Data, and Space***, published by Pearson Scott Foresman, uses a student-centered approach focused on conceptual understanding, rather than on students’ abilities to answer problems correctly.
- ***Math Expressions***, published by Houghton Mifflin Harcourt, blends student-centered and teacher-directed approaches to mathematics. Students question and discuss mathematics, but are explicitly taught effective procedures.

¹ Given the number of schools and classrooms included in the study, the statistical power benefits of pre- and post-testing more than 10 students per classroom are minimal, though the costs are significant because the study used an individually-administered assessment.

- *Saxon Math*, published by Houghton Mifflin Harcourt, is a scripted curriculum that blends teacher-directed instruction of new material with daily distributed practice of previously learned concepts and procedures. Students are explicitly taught procedures and strategies.
- *Scott Foresman-Addison Wesley Mathematics*, published by Pearson Scott Foresman, is a basal curriculum that uses a teacher-directed approach. The curriculum is based on a consistent daily lesson structure which includes direct instruction, hands-on exploration, the use of questioning, and practice of new skills.

Publishers provided curriculum materials and training to teachers in the summer prior to the first day of school; additional follow-up training was provided by the publishers to study teachers during the school year.

Research Design:

About a quarter of the study's schools were randomly assigned to each of the study's four curricula. Random assignment of curricula to schools was conducted separately for each participating district, which established an experiment in each study district. A stratified random assignment approach was used, in order to allocate a similar numbers and types of schools, teachers, and students to each curriculum. In particular, the procedure divided schools in each district into strata, where each stratum contained schools with similar baseline characteristics. Random assignment of curricula to schools then took place within each stratum. The procedure helped minimize chance differences in school characteristics and sample sizes across curriculum groups, thus helping increase the design's statistical power and face validity.

Data Collection and Analysis:

The study team collected all data necessary to evaluate the four curricula. To measure the achievement effects of the curricula, the study team tested students at the beginning and end of the school year using the math assessment developed for the Early Childhood Longitudinal Study (ECLS-K) (West et al. 2000). The ECLS-K assessment is a nationally normed test designed to measure achievement gains both within and across elementary grades. The first- and second-grade results are based on students who were tested in both the fall and spring in those respective grades.

To help interpret the measured achievement effects, teachers completed surveys about curriculum implementation, and the study team observed each first- and second-grade classroom once during the school year. Together, the survey and observation data are useful for assessing teacher participation in curriculum training, use of the assigned curriculum, and supplementation of the assigned curriculum with other materials. The data also were useful for assessing adherence to each curriculum's specific features and for examining curriculum-group differences in teaching approaches and practices that could be measured consistently across the curricula.

The relative effects of the curricula were estimated using a three-level hierarchical linear model (HLM), which incorporates the nested structure of the data when calculating the statistical significance of the results.² To help offset the losses in precision from clustering, the HLM

² Clustering occurs at the school level because, if random assignment were repeated, a different set of classrooms would be assigned to the study's curricula. Clustering also occurs at the classroom level because, as only a sample of students in each classroom was tested, a different set of students would be tested if sampling were repeated.

included (as covariates) baseline student, teacher, and school measures related to student achievement. Most importantly, the model included student math achievement at the beginning of the school year as a covariate. Only results that are statistically significant at the 5 percent level of confidence are discussed.

Findings / Results:

In terms of student math achievement, the curriculum used by the study schools mattered (please insert Figure 1 here). In first grade classrooms, average math achievement of Math Expressions students was 0.11 standard deviations higher than that of both Investigations and SFAW students; in second grade classrooms average math achievement of Math Expressions and Saxon students was 0.12 and 0.17 standard deviations higher than that of SFAW students, respectively. None of the other curriculum differentials are statistically significant.³

These findings are based on statistical tests that have not been adjusted for the six unique pair-wise curriculum comparisons that can be made with the four curricula. Results based on statistical tests that have been adjusted (using the Tukey-Kramer method) for multiple comparisons indicate that only the second-grade Saxon-SFAW differential of 0.17 standard deviations is statistically significant.⁴

The study also examined the relative effects of the curricula for subgroups; the curriculum used in different contexts also mattered, and some of these findings are consistent with findings based on all students whereas others are not.

Last, the study used an instrumental variables (IV) approach to examine what accounts for the relative curriculum effects and found evidence of mediation for some of the implementation measures that differ across the curriculum groups. However, the evidence of mediation was inconsistent across the curriculum-pair differentials that are statistically significant. For example, according to the IV approach, math instructional time mediates three of the four curriculum-pair differentials that are statistically significant (according to unadjusted statistical tests) across the two grade levels, but it does not mediate the fourth differential.

Conclusions:

As math skills have grown in importance in the workplace, so has the debate over how best to improve students' math skills. This study is an important step toward resolving that debate because it shows that four widely-used curricula have differential effects on student math achievement. These differential effects, which were measured after just one year of curriculum implementation, suggest that districts should carefully choose their math curriculum.

A future report will examine second-year effects of the curricula in study schools that participated in the study for a second year.

³ The relative effects of the curricula reflect all differences between the curricula, including differences in teacher training, instructional strategies, content coverage, and curriculum materials.

⁴ A large literature that considers the issue of multiple comparison adjustments, but, to our knowledge, there is no consensus about whether statistical tests should or should not be adjusted (see, for example, Saville 1990 and Westfall et al. 1999). For this reason, both sets of results are presented.

Appendix A. References

Denton, K., and Jerry West. (2002). *Children's Reading and Mathematics Achievement in Kindergarten and First Grade*. Publication No. NCES 2002-125. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hiebert, J.C., and D.A. Grouws. (2007). "The Effects of Classroom Mathematics Teaching on Students' Learning. In *Second Handbook of Research on Mathematics Teaching and Learning*, edited by F.K. Lester, Jr. (vol. 1, pp. 371–404). New York: Information Age Publishing.

National Center for Education Statistics. (2009). *The Nation's Report Card: Mathematics 2009*. Publication No. NCES-2010-451. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Mathematics Advisory Panel. (2008). *Foundations for success: the final report of the national mathematics advisory panel*. Washington, DC: U.S. Department of Education.

National Research Council. (2004). *On evaluating curricular effectiveness: judging the quality of K-12 mathematics evaluations*. Washington, DC: National Academies Press.

Rathburn, A., and J. West. (2004). *From Kindergarten Through Third Grade: Children's Beginning School Experiences*. Publication no. NCES-2004-007. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

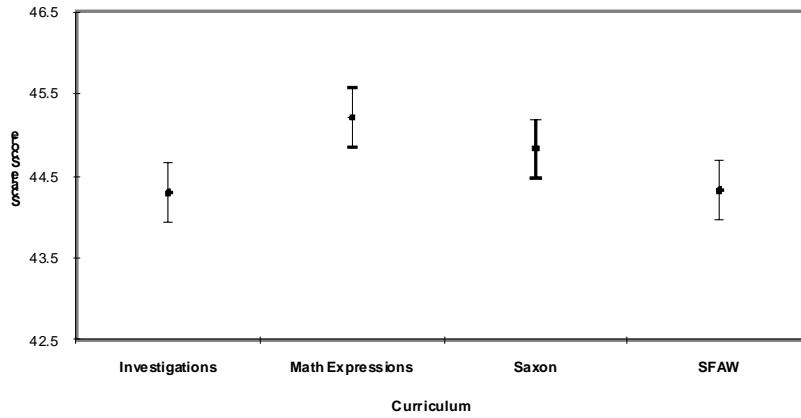
Resnick, Robert M., Glenn Sanislo, and Stephanie Oda. (2010). *The Complete K–12 Report: Market Facts and Segment Analyses*. Rockaway Park, NY: Education Market Research.

West, Jerry, Kristin Denton, and Elvira Germino-Hausken. (2000). *America's Kindergartners*. Publication No. NCES 2000-070. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

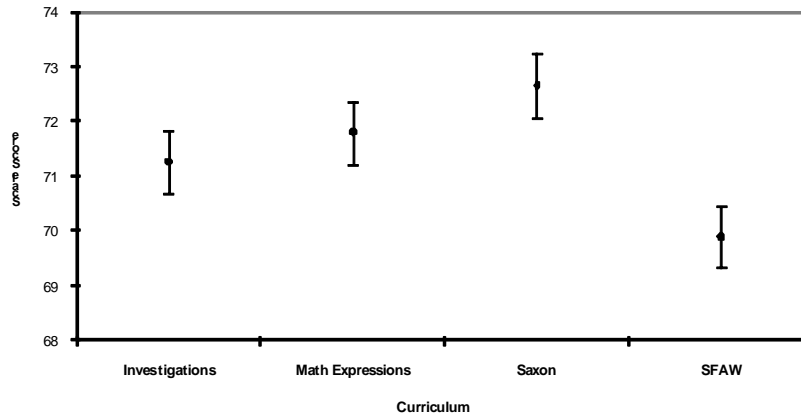
Appendix B. Figure 1

AVERAGE HLM-ADJUSTED SPRING STUDENT MATH SCORE WITH CONFIDENCE INTERVAL,
BY GRADE AND CURRICULUM

First-Grade Students



Second-Grade Students



Note: The dots in each symbol represent the average HLM-adjusted spring student math score for each curriculum, and the bars that extend from each dot represent the 95 percent confidence interval around each average. Curricula with non-overlapping confidence intervals have significantly different average scores at the 5 percent level. Each curriculum was randomly assigned to about 27 schools, 116 classrooms, and 1,180 students for the first-grade analysis, and to about 18 schools, 82 classrooms, and 835 students for the second-grade analysis.