# How Four Elementary Math Curricula Perform Among Different Types of Teachers and Classrooms

Roberto Agodini
Barbara Harris
Mathematica Policy Research

December 2014

## Abstract

Research shows that some elementary math curricula are more effective than others at increasing student achievement. Most studies in this research base typically included teachers and classrooms that differ, so the results represent *average* curriculum effects, which raises an important question: Are curricula that are effective on average also effective among different types of teachers and classrooms? We examine whether curriculum effects are moderated by three characteristics that influence curriculum implementation and, therefore, may influence effects: (a) teachers' knowledge, (b) teachers' attitude toward math instruction, and (c) the extent to which teachers need to differentiate instruction in their classrooms. We examine these moderators for four elementary math curricula that use different pedagogical approaches, and for first and second grade achievement. We find that the two curricula which are more effective on average also are either as effective, or more effective than the other two across all the contexts examined.

# How Four Elementary Math Curricula Perform Among Different Types of Teachers and Classrooms

The math curricula used in elementary schools across the United States vary in their approaches to teaching and learning (Stein, Remillard, & Smith, 2007).[1] For example, curricula differ in their treatment of content and learning experiences; some curricula place higher cognitive demands on students by having them figure out important mathematical ideas that are not immediately apparent (Stein & Kim, 2009). In addition, they differ in the sequencing of content and the extent of review of previously learned skills (Remillard, Harris, & Agodini, 2014). Finally, some programs are clearer about the objectives of the lessons and provide more teacher supports for implementing them, including more pedagogical guidance (Stein & Kaufman, 2010).

Research shows that math curricula (and their associated textbooks) have a strong influence on what students are taught in mathematics (Mark, Spencer, Zeringue, & Schwinden, 2010; Grouws, Smith, & Sztajn, 2004), and different curricula can have different effects on student achievement. The U.S. Department of Education's What Works Clearinghouse (2014) reviewed effectiveness studies of several elementary math curricula; looking across those reviews suggests that some curricula are more effective at raising student achievement than others.

We recently completed a large-scale (110 schools), randomized controlled trial (RCT) of four widely used elementary math curricula that adds to the research base about curriculum effects on student achievement. Looking across the study's results, two of the curricula were more effective than the other two at improving math achievement of first and second-grade students (Agodini & Harris 2010; Agodini, Harris, Thomas, Murphy, & Gallagher 2010; Agodini, Harris, Seftor, Remillard, & Thomas, 2013).

The two more effective curricula in our RCT are based on different pedagogical approaches, which raise the issue of *why* two different pedagogical approaches were equally effective. Two of the study's curricula were developed with support from the National Science Foundation (NSF) and are often categorized as standards-based curricula that use a constructivist approach to instruction, whereas the other two are often categorized as conventional curricula that use a more explicit approach to instruction (Stein et al., 2007). The two more effective curricula in the RCT included one standards-based and one conventional curriculum.

At least three follow-up analyses could help us better understand why the RCT found that two curricula with different pedagogical approaches were most effective. One analysis would examine whether there are common characteristics among the two outperforming curricula that differ from the other two. Another analysis would explore whether the findings are consistent among students in other grades than those examined in the RCT. The final analysis would consider whether the findings are robust across different types of teachers and classrooms. Our RCT (like many effectiveness studies in education) included teachers and classrooms that differ, so the results represent *average* curriculum effects. The last follow-up analysis would help us understand whether the pedagogical approaches that underlie the two more effective curricula are robust across the different types of teachers and classrooms included in our RCT.

In this paper, we use the RCT's data to address the third issue. Specifically, we examine the moderating effects of the following teacher and classroom characteristics: (a) teachers' knowledge, (b) teachers' attitudes toward math instruction, and (c) the extent to which teachers need to differentiate instruction in their classrooms, as measured by the range of student ability in the classroom. We focus on these characteristics because they are available in the RCT's data and, more importantly, because other research shows they influence curriculum implementation and, therefore, may influence effects.

**Rationale**

Research shows that curriculum implementation varies across different types of teachers. These studies have found that teachers' knowledge of math content and pedagogy, their orientation toward curriculum, and their professional identity influence curriculum implementation (Stein et al., 2007; Stein & Kaufman, 2010). These studies show that organizational characteristics also matter, including classroom structures, principal support for a curriculum, and extent of a school's/district's professional community (Stein et al., 2007). Consistent with these findings, other research shows that teachers' ability to implement a curriculum and buy-in for a new proposed curriculum factor into districts' adoption decisions (Zeringue, Spencer, Mark, & Schwinden, 2010).

The findings of this research beg an important question: Do the characteristics that influence curriculum implementation moderate curriculum effects on student achievement? To our knowledge, no rigorous evidence exists on whether the fit between these characteristics and curriculum matters for student achievement.

The teacher and classroom characteristics we examine are important to investigate for two reasons. First, research shows that teachers and classrooms differ along these dimensions, so it is important to understand which curricula work best in which setting. For example, Hill and Lubienski (2007) found that teacher knowledge varies significantly across schools. In addition, when recruiting districts to participate in the RCT mentioned above, we learned that teachers often have strong attitudes about how to teach mathematics. In fact, some districts were not willing to participate in the RCT because they felt that at least one of the study's curricula should not be the basis for math instruction. Finally, Loveless (2013) showed that nearly two-thirds of fourth grade teachers in 2011 used ability grouping for math instruction—a practice that has increased significantly in recent years.

Second, districts are collecting systematic information on these characteristics, so they have information that is useful for selecting curricula that may work best in their settings. For example, new federal policies such as Race to the Top, created a national push for new teacher evaluations that provide schools and districts with systematic, comprehensive information on teachers' knowledge, attitudes, and practices. Information on the range of student abilities has been readily available for several years through annual student assessments.

We shed light on the importance of teacher and classroom characteristics for curriculum effects by answering the following question:

- Do three of the teacher and classroom characteristics that influence curriculum implementation—namely, teachers' knowledge, their attitudes about instruction, and the range of student ability in the classroom—moderate curriculum effects?

## Method

### Participants

A total of 110 schools from 12 districts participated in the RCT study in either the 2006-2007 or the 2007-2008 school year. Though not a representative sample of all elementary schools in the United States, the study schools are dispersed geographically and located in areas with various levels of urbanicity. A larger percentage also are schoolwide Title I eligible than the average U.S. elementary school (57 versus 44 percent).

At the outset of the RCT, each of the 110 schools was randomly assigned to implement one of four curricula: (a) Investigations in Number, Data, and Space (Investigations); (b) Math Expressions; (c) Saxon Math (Saxon); and (d) Scott Foresman-Addison Wesley Mathematics (SFAW). These curricula are widely used in early elementary grades and differ in their approaches to mathematics instruction. Investigations and Math Expressions are often categorized as standards-based curricula, whereas and Saxon and SFAW are often categorized as conventional curricula (Stein et al., 2007).

The random assignment was conducted separately in each district, after all teacher consent forms for all participating schools in a district were received. Within each participating school, all teachers in the participating grade levels (first and/or second grades) participated in the study, and used their school's assigned curriculum with all of their students in the classroom.

**Materials**

**Student Achievement.** To measure the effects of the curricula on student achievement, the RCT administered at the beginning and end of the school year the math assessment from the Early Childhood Longitudinal Study-Kindergarten (ECLS-K) Class of 1998-1999 study. The ECLS-K assessment is an individually administered, adaptive, and nationally normed test that measures student achievement both within and across grades (Rock & Pollack, 2002).

**Teacher Knowledge.** Before the start of the initial curriculum training sessions, the RCT team assessed teachers' mathematical knowledge for teaching (MKT) using an assessment developed by researchers at the University of Michigan. It included items about teacher pedagogical content knowledge in two domains: knowledge of mathematics for teaching, and knowledge of how students learn mathematics. Items focused on numbers, operations, and patterns; functions; and algebra—the three content areas most frequently covered in the elementary grades. Hill, Schilling, and Ball (2004) provide details about the assessment's development process, and Hill and Ball (2004) provide the sample items. Item response theory (IRT) techniques were used to calculate the overall teacher score, as recommended by the test developers. The reliability of the teacher test score for the study's sample equals 0.75.

**Teacher Attitudes.** In the fall of the first year of the RCT, a survey was administered that asked teachers to report the extent to which they agreed with 11 statements designed to measure different views about how students should be taught math. Table 1 lists the teacher responses to each of the 11 items. By design, about half the statements measured teachers' affinity with practices that are consistent with a "constructivist" approach to instruction; the rest measured teachers' affinity with practices that are consistent with an "explicit" approach. Teachers rated each statement on a four-point scale (strongly disagree, disagree, agree, and strongly agree). Responses were aggregated across the strongly disagree and disagree categories in Table 1 to protect respondent confidentiality.

We created two scales from the survey data labeled: (a) constructivist instruction, and (b) explicit instruction. The constructivist instruction scale includes the first five items in Table 1, the explicit instruction scale the remaining six. The reliability of the constructivist and explicit instruction scales equals 0.78 and 0.76, respectively.

**Range of Student Ability.** Most educators in the RCT reported having a range of student abilities within each classroom. One of the questions we often encountered during district recruitment for the RCT was whether the study's curricula were appropriate for diverse learners. Each of the publishers reported that the materials were designed to support differentiated instruction, but some educators questioned whether each curriculum was equally supportive of differentiated instruction. Therefore, we examine whether curriculum effects are moderated by teachers' need to differentiate instruction in their classrooms, as measured by the range of student ability. To measure the range of student ability, we calculated the classroom-level standard deviation of the pre-test scores on the ECLS-K math assessment.

**Sample Sizes and Response Rates**

Across the 110 study schools 789 first- and second-grade teachers participated in the study. In each classroom, an average of about 12 students was randomly selected for testing at the beginning of the school year, for a total of 9,712 students. Of this sample, 8,060 were tested at both the beginning and the end of the school year, for a response rate of 83 percent, ranging from 82 to 84 percent across the curriculum groups.

Teacher assessments were completed by 95 percent of the study teachers, and the responses rates across the curriculum groups were similar, ranging from 93 to 96 percent (Agodini et al., 2010). The survey-based measures of teachers' attitudes toward math instruction were available for 93 percent of study teachers, and the response rates were similar across curriculum groups, ranging from 91 to 96 percent.

**Design**

To examine whether the teacher/classroom characteristics moderate curriculum effects, we estimated a separate statistical model of student achievement for each characteristic, where each model includes an interaction term for the characteristic and curriculum. The coefficient on the interaction term indicates whether the curricula differentially impact achievement of students in different classroom environments.

Specifically, we estimated a separate hierarchical linear model (HLM) for each teacher/classroom characteristic that includes the following three equations for student, classroom, and school data:

$$POST_{ijk} = \alpha_{0jk} + \alpha_{1jk} PRE_{ijk} + \varepsilon_{ijk} \qquad (1)$$

$$\alpha_{0jk} = \beta_{0k} + \beta_{1k}T_{jk} + \mu_{jk} \tag{2}$$

$$\beta_{0k} = \delta_1 + \sum_{c=2}^{4}\delta_c C_k + \sum_{c=2}^{4}\lambda_c C_k \times T_{jk} + \nu_k \tag{3}$$

The first (student-level) equation assumes that spring math achievement can differ across students—that is, $POST_{ijk}$ equals spring math achievement of student *i* in classroom *j* in school *k*, and $\alpha_{0jk}$ equals average spring math achievement of all students in classroom *j* adjusted for average math achievement at the beginning of the school year as represented by $PRE_{ijk}$. The second (classroom-level) equation assumes that average adjusted spring classroom achievement from the first equation ($\alpha_{0jk}$) can differ across classrooms—that is, $\beta_{0k}$ equals average spring math achievement in school *k* adjusted for average student characteristics in the first equation, and average classroom characteristics (including one of the teacher/classroom characteristics that are the focus of this study) as represented by $T_{jk}$ in the second equation. The third (school-level) equation assumes that average adjusted spring school achievement from the second equation ($\beta_{0k}$) can differ across schools, and may depend on the curriculum randomly assigned to schools ($C_k$) and the interaction between curriculum and the teacher/classroom characteristic included in the second equation ($C_k$*$T_{jk}$).[2]

According to our three-level HLM, the effect of curriculum $C_k$ relative to curriculum 1 (which is not included in the model and therefore serves as the curriculum to which the included curricula are compared) is:

$$\frac{\partial POST_{ijk}}{\partial C_k} = \delta_k + \lambda_k T \tag{4}$$

In other words, the model assumes that the effect of curriculum $C_k$ relative to curriculum 1 includes a main effect ($\delta_c$) and an interaction effect ($\lambda_c$) that depends on the value of the teacher/classroom characteristic *T* included in the model.

For each specification of the HLM, we use the end-of-year score as the measure of $POST_{ijk}$ and the beginning-of-year score as the measure of $PRE_{ijk}$. We estimate one specification of the HLM using the overall score teachers received on the MKT assessment as the measure of $T_{jk}$, another specification using the constructivist instruction scale as the measure of $T_{jk}$, a third specification using the explicit instruction scale as the measure of $T_{jk}$, and a fourth specification using the classroom-level standard deviation of the pre-test scores as the measure of $T_{jk}$. Each HLM was estimated with the 8,060 students using a student-level weight that sums to the number of students in each classroom that were eligible for testing at the beginning of the school year.[3]

To understand the influence of each teacher/classroom characteristic, we compare curriculum effects for the average teacher (that is, when equation 4 is evaluated at the average value of $T$) with effects for a teacher who is one standard deviation above the mean (that is, the value of equation 4 when $T$ equals the mean value of $T$ plus the standard deviation of $T$).

## Results

Tables 2 through 5 present results from the HLMs that separately examine each of the teacher/classroom characteristics we hypothesize influence curriculum effects. Each table presents curriculum effects for two types of teachers/classrooms: (a) those with the average value for the characteristic included in the HLM, and (b) those whose characteristic is one standard deviation above the mean. These two results are presented for each of the six *unique* pair-wise curriculum comparisons that can be made with the study's four curricula: Investigations relative to Math Expressions; Investigations relative to Saxon; Investigations relative to SFAW; Math Expressions relative to Saxon; Math Expressions relative to SFAW; and Saxon relative to

SFAW. End of year student achievement (the dependent variable) is in standard deviation units, so all results indicate effect sizes.

For each of the six curriculum differentials in a table, we compare the results for the two types of teachers and conclude there is evidence of moderation if one of the following two conditions holds. One condition requires that the curriculum effects for both types of teachers/classrooms are statistically significant and the two effects differ by at least 0.03 standard deviations. The other condition requires that one effect is statistically significant and at least 0.03 standard deviations, whereas the other effect is not statistically significant. We focus on a difference of at least 0.03 standard deviations because it is equivalent to a 1-percentile point change in achievement for the average student.

We find that each of the teacher/classroom characteristics influences at least one of the curriculum effects and most often the effect of Investigations, sometimes in a negative way and sometimes in a positive way. Looking across Tables 2 through 5, there are seven situations where a curriculum differential is moderated by the teacher/classroom characteristics examined. For example, as shown in Table 2, two curriculum differentials—Investigations relative to Math Expressions and Investigations relative to Saxon—are significantly different across teachers with average and above average MKT scores. In five of the seven situations (over two-thirds of the time) where there is evidence of moderation, the moderated effect involves Investigations. In the other two situations, the effect of Saxon relative to SFAW is moderated. In the discussion section below, we describe the results in more detail and provide potential explanations for our findings.

# Discussion

## Interpreting the Curriculum Moderators

In this section, we discuss the findings for each of the three moderators we examined. In the next section, we discuss our main conclusions and the policy implications of our findings. In the final section we provide suggestions for future research that would further illuminate our understanding of the moderators of curriculum effects.

**Teacher Knowledge.** Among teachers with average MKT, the effect of Investigations is not significantly different from Math Expressions and Saxon, but Investigations is less effective than Math Expressions and Saxon among teachers with above-average MKT (Table 2). This result suggests that teachers' MKT moderates Investigations' effect so strongly that, among teachers with average MKT, Investigations is not significantly different from the most effective curricula (Math Expressions and Saxon), but Investigations is one of the least effective curricula (along with SFAW) among teachers with above-average MKT. The effect of SFAW relative to the other three curricula is unaffected by increases in teacher MKT.

This Investigations result is surprising because this curriculum is largely focused on developing students' conceptual understanding of mathematics, so it seems that its effect relative to the other curricula should increase as teachers' knowledge of how students learn mathematics increases. This unexpected finding may be explained by the findings of other research, which shows that Investigations is more challenging to implement than other curricula (Stein & Kaufman, 2010). While more knowledgeable teachers may make better use of Investigations than teachers with average knowledge, perhaps more knowledgeable teachers make even better use of Math Expressions and Saxon because those two curricula may be easier to implement than Investigations.

**Teacher Attitudes.** The average study teacher agrees fairly strongly with constructivist instruction. The average value of the five items that constitute the constructivist scale indicates that the average study teacher's view toward this instructional approach lies between "agree" and "strongly agree."

The effect of Investigations relative to Math Expressions is lower for teachers who agree more strongly with constructivist instruction than the average teacher (Table 3). This finding is surprising, since Investigations is the most constructivist of the study curricula and, therefore, one might think it would be better aligned with teachers who strongly believe in a constructivist approach.

However, if we view instruction along a continuum with constructivist approaches at one end and explicit approaches at the other, this findings might suggest that the ideal instructional approach in first and second grades is somewhere in the middle and includes both constructivist and explicit approaches. Because Investigations is the most constructivist of the study curricula, perhaps when it is used by teachers who strongly believe in such approaches, the classroom environment moves closer to the constructivist end of the continuum and farther away from the ideal balance, and thus becomes less effective.

There is one finding that appears inconsistent with this hypothesis, though we consider the contradictory evidence weak because the evidence does not meet our requirements that the results for the average and above-average teacher differ by at least 0.03 standard deviations. Among teachers with the average view toward constructivist instruction, Investigations is less effective than Saxon. Among teachers who agree more strongly with constructivist instruction, the Investigations-Saxon differential is not statistically significant at the 5 percent level but is significant at the 10 percent level, and indicates that Investigations is less effective than Saxon by a similar amount as what we observed for the average teacher.

Turning to the moderating effect of teachers' views toward explicit instruction (Table 4), these results also could be an indication that the ideal instructional approach in first and second grades involves a balance of constructivist and explicit instruction. The mean of the six items that constitute the explicit instruction scale indicates that the average study teacher "agrees" with this instructional approach. Among teachers with the average view toward explicit instruction, the results suggest that Investigations outperforms SFAW by 0.06 standard deviations—we say "suggest" because this result is not statistically significant at the 5 percent level, but is significant at the 10 percent level. Among teachers who agree more strongly with explicit instruction, the benefit of using Investigations over SFAW is even larger at 0.14 standard deviations. This finding is surprising, since one might expect that teachers with a strong affinity toward explicit instruction would do better with SFAW than Investigations because SFAW is the more explicit of the two curricula. However, these data indicate the opposite. Thus, when Investigations (a heavily constructivist curriculum) is used by teachers with a strong affinity for explicit instruction, the enacted curriculum may be closer to the ideal balance of constructivist and explicit instruction in first and second grades.

**Range of Classroom Ability.** In classrooms with the average range of student ability (Table 5), Investigations is less effective than Math Expressions, but the achievement difference between these curricula is not statistically significant in classrooms where the range of ability is greater. Thus, the benefit of using Math Expressions over Investigations reduces to zero as a teachers' need to differentiate instruction increases.

This result suggests Investigations might be easier to use in first and second grade classrooms when teachers have a wider range of students that need differentiated instruction—a result that is not necessarily surprising given the nature of the Investigations materials. For example, a common Investigations activity is to have students show (or write) all of the ways to represent a particular number. Teachers could easily tailor this activity to meet a wide range of student ability by having some students represent a small or easy number, while simultaneously having other students represent a much larger or complicated number.

Classroom ability also moderates the Saxon-SFAW differential, but in the opposite way than what we observe for the Investigations-Math Expressions differential. In classrooms with the average range of student ability, Saxon is more effective than SFAW; Saxon is even more effective than SFAW in classrooms where the range is wider. This finding suggests that the benefit of using Saxon over SFAW increases as the need to differentiate instruction increases. This finding is surprising, since SFAW is designed to support differentiated instruction. However, differentiated practice in SFAW is largely implemented through the use of differentiated student worksheets. Perhaps study teachers did not use them to their full potential, since it would require teachers to prepare materials from a variety of different workbooks.

**Conclusions and Policy Implications**

The goal of this study was to examine whether curriculum effects are moderated by three teacher and classroom characteristics: (a) teachers' knowledge, (b) teachers' attitudes toward math instruction, and (c) the extent to which teachers need to differentiate instruction in their classrooms, as measured by the range of student ability in the classroom. Examining the moderating effect of these teacher and classroom characteristics is important because research shows that they influence curriculum implementation and, therefore, may influence curriculum effects. We examined how these factors moderate the effects of two standards-based curricula (Investigations and Math Expressions) and two conventional curricula (Saxon and SFAW) in first and second grades.

We found that the effects of one of the standards-based and one of the conventional curricula (Math Expressions and Saxon, respectively) are more robust across the teacher/classroom contexts examined, than the other standards-based and the other conventional curriculum (Investigations and SFAW, respectively). In addition, across all the teacher and classroom contexts examined, Math Expressions and Saxon (the two more robust curricula) are either as effective, or more effective than Investigations and SFAW (the two less robust curricula). There are situations where the effectiveness of Investigations and SFAW is greater than what is observed for the average teacher or classroom. However, in all of these situations, these curricula are, at best, equally effective as Math Expressions and Saxon.

Districts may find these results useful for at least three reasons. First, the results suggest that curriculum is an effective lever for improving student math achievement in first and second grades, but some curricula are a better choice than others across the different contexts we examined. Specifically, if districts are thinking about selecting curricula such as Investigations or SFAW for first and second grades, our findings suggest that they should consider their teachers' knowledge, attitudes about instruction, and need to differentiate instruction because these contextual factors influence the effects of these curricula. For example, among teachers with average knowledge of how students learn math, Investigations is not significantly different from the most effective curricula (Math Expressions and Saxon), but is one of the least effective curricula (along with SFAW) among teachers with above-average knowledge. In all other contexts examined, Math Expressions or Saxon is also a better choice than Investigations.

Second, districts may be interested to know that the more effective curricula differ in their approaches to instruction and learning, so educators can choose a curriculum that best suits their teaching style. As mentioned above, Math Expressions is often categorized as a standards-based curriculum, whereas and Saxon is often categorized as a conventional curriculum (Stein et al., 2007). Since some districts feel strongly that one of these curriculum types should (or should not) be the basis for instruction, our results suggest that districts could adopt an effective curriculum that matches their preferred instructional approach.

Finally, districts may find the results useful for understanding whether curriculum effects can be maximized by enhancing teachers' knowledge of how students learn mathematics or altering the way students are assigned to classrooms based on ability in first and second grades. For example, our results indicate that investing in professional development focused on teacher knowledge would increase the benefit of using Math Expressions and Saxon over Investigations in first and second grades (Table 2). Research shows that such teacher knowledge can be improved through professional development (Hill & Ball, 2004), suggesting that careful curriculum selection combined with professional development could maximize the benefit of using curriculum as a lever to improve student achievement. Our results also indicate that less use of ability grouping in classroom assignments would increase the benefit of using Saxon over SFAW (Table 5) in first and second grades, suggesting districts may wish to take into account their policies related to ability grouping when selecting a math curriculum.

# References


Agodini, R., & Harris, B. (2010). An experimental evaluation of four elementary school math curricula. *Journal of Research on Educational Effectiveness, 3*(3), 199-253.

Agodini, R., Harris, B., Seftor, N., Remillard, J., & Thomas, M. (2013, September). After two years, three elementary math curricula outperform a fourth. (NCEE 2013-4019.) Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Agodini, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010, October). Achievement effects of four early elementary school math curricula: Findings for first and second graders. (NCEE 2011-4001.) Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Agodini, R., Deke, J., Atkins-Burnett, S., Harris, B., & Murphy, R. (2008, January). Design for the evaluation of early elementary school mathematics curricula. Princeton, NJ: Mathematica Policy Research.

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *Elementary School Journal, 108*(2), 115–130.

Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102*(2), 294–343.

Grouws, D.A., Smith, M.S., & Sztajn, P. (2004). The preparation and teaching practices of U.S. mathematics teachers: Grades 4 and 8. In P. Kloosterman & F. Lester (Eds). *The 1990 through 2000 mathematics assessments of the National Assessment of Educational Progress: Results and interpretations* (pp. 221-269). Reston, VA: National Council of Teachers of Mathematics.

Guarino, C., Dieterle, S. G., Bargagliotti, A. E., & Mason, W. M. (2013). What can we learn about effective early mathematics teaching? A framework for estimating causal effects using longitudinal survey data. *Journal of Research on Educational Effectiveness, 6*(2), 164–198.

Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education, 35*(5), 330–351.

Hill, H. C., & Lubienski, S. T. (2007). Teachers' mathematics knowledge for teaching and school context: A study of California teachers. *Educational Policy, 21*(5), 747-768.

Hill, H., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, *105*(1), 11–30.

Le, V., Stecher, B., Lockwood, J. R., Hamilton, L., Robyn, A., Williams, V., & Klein, S. P. (2006). *Improving mathematics and science education: A longitudinal investigation between reform-oriented instruction and student achievement.* Santa Monica, CA: RAND Corporation.

Loveless, T. (2013). The 2013 Brown Center report on American education: How well are American students learning? Washington, DC: The Brown Center on Education Policy, The Brookings Institution.

Mark, J., Spencer, D., Zeringue, J., & Schwinden, K. (2010). *How do districts choose mathematics textbooks?* In B. Reys and R. Reys (eds.), NCTM 2010 Yearbook, The K-12 Mathematics Curriculum: Issues, Trends, and Future Directions. Reston, VA: National Council of Teachers of Mathematics.

Rock, D. A., & Pollack, J. M. (2002) Early childhood longitudinal study—Kindergarten class of 1998–99 (ECLS-K), Psychometric report for kindergarten through first grade. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, *18*(6), 565–579.

Stein, M. K., & Kaufman, J. H. (2010). Selecting and supporting the use of mathematics curricula at scale. *American Educational Research Journal, 47*(3), 663-693.

Stein, M. K., & Kim, G. (2009). The role of mathematics curriculum materials in large-scale urban reform: An analysis of demands and opportunities for teacher learning. In J. T. Remillard, B. A. Herbel-Eisenmann, & G. M. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 37–55). New York, NY: Routledge.

Stein, M. K., Remillard, J., & Smith, M. S. (2007). How curriculum influences student learning. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 319–369). New York, NY: Information Age Publishing.

Tarr, J. E., Reys, R. E., Reys, B. J., Chávez, O., Shih, J., & Osterlind, S. J. (2008) The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education*, 39(3), 247–280.

What Works Clearinghouse. (2014). Website. Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/

Zeringue, J., Spencer, D., Mark, J., & Schwinden, K. (2010). *Influences on Mathematics Textbook Selection: What Really Matters?* Presented at the NCTM Research Pre-Session, San Diego, CA.

## Notes

This research was supported by the Smith Richardson Foundation. This draft of the paper was shortened to fit within NCTM's page limit. A more complete version is available from the authors.

[1] In this paper, we use the term curriculum to reference the complete package of materials offered by publishers, which includes pedagogical guidance, the content to be covered, and the student and teacher materials that support instruction throughout a school year.

[2] The $C_k$ terms equal one for schools assigned to curriculum $k$ and zero otherwise. Interacting curriculum and the teacher/classroom characteristic is a cross-level interaction.

[3] The weight was not adjusted for nonresponse to testing (mainly because of parental nonconsent and within-school-year attrition), because a nonresponse analysis showed that none of the available baseline characteristics were related to nonresponse.

Table 1
Frequency of teacher responses to specific instructional practices, by scale (percentages)

| Extent to which teachers agree with the following statements | Strongly Disagree or Disagree | Agree | Strongly agree |
|---|---|---|---|
| Constructivist approach to instruction scale | | | |
| Believe that formative assessments are important | 2.2 | 60.1 | 37.8 |
| Believe that students learn from one another when they work together | 1.0 | 51.1 | 48.0 |
| Believe that students learn best when they share their reasoning with one another | 5.8 | 59.8 | 34.4 |
| Believe that it is important that teachers observe and listen to how students think | — | 44.7 | 55.3 |
| Believe that a math concept is learned best if grounded in real life when introduced | 5.0 | 58.9 | 36.1 |
| Explicit approach to instruction scale | | | |
| Believe that when students ask questions, teacher should provide a thorough explanation | 26.0 | 49.7 | 24.3 |
| Believe that it is important that students not have misconceptions at the end of the lesson | 11.9 | 60.2 | 27.8 |
| Believe that students should demonstrate mastery of one concept before moving to the next | 43.8 | 42.1 | 14.1 |
| Believe that teachers should emphasize computational skills | 15.0 | 67.9 | 17.1 |
| Believe that teachers should clearly model solution approaches for new types of problems | 12.9 | 44.0 | 43.1 |
| Believe that students should primarily work individually | 66.2 | 27.7 | 6.1 |

Note: The statistics are based on 789 teachers.

— Value suppressed to protect respondent confidentiality.

Table 2
Curriculum effects by teacher knowledge

| Curriculum effects | Effect size | *p*-Value |
|---|---|---|
| *Investigations relative to Math Expressions for teachers with:* | | |
| Average MKT Score | -0.08 | 0.15 |
| MKT Score One Standard Deviation Above the Mean | -0.16* | 0.00 |
| *Investigations relative to Saxon for teachers with:* | | |
| Average MKT Score | -0.06 | 0.44 |
| MKT Score One Standard Deviation Above the Mean | -0.13* | 0.01 |
| *Investigations relative to SFAW for teachers with:* | | |
| Average MKT Score | 0.04 | 0.29 |
| MKT Score One Standard Deviation Above the Mean | -0.03 | 0.43 |
| *Math Expressions relative to Saxon for teachers with:* | | |
| Average MKT Score | 0.02 | 0.74 |
| MKT Score One Standard Deviation Above the Mean | 0.03 | 0.49 |
| *Math Expressions relative to SFAW for teachers with:* | | |
| Average MKT Score | 0.11* | 0.00 |
| MKT Score One Standard Deviation Above the Mean | 0.13* | 0.01 |
| *Saxon relative to SFAW for teachers with:* | | |
| Average MKT Score | 0.10* | 0.01 |
| MKT Score One Standard Deviation Above the Mean | 0.09* | 0.04 |

*Statistically significant at the 5 percent level.

Table 3
Curriculum effects by teacher constructivist views

| Curriculum effects | Effect size | *p*-Value |
|---|---|---|
| Investigations relative to Math Expressions for teachers with: | | |
| Average View Toward Constructivist Instruction | -0.08* | 0.03 |
| View One Standard Deviation Above the Mean | -0.13* | 0.01 |
| Investigations relative to Saxon for teachers with: | | |
| Average View Toward Constructivist Instruction | -0.07* | 0.05 |
| View One Standard Deviation Above the Mean | -0.09 | 0.06 |
| Investigations relative to SFAW for teachers with: | | |
| Average View Toward Constructivist Instruction | 0.03 | 0.44 |
| View One Standard Deviation Above the Mean | -0.02 | 0.69 |
| Math Expressions relative to Saxon for teachers with: | | |
| Average View Toward Constructivist Instruction | 0.01 | 0.83 |
| View One Standard Deviation Above the Mean | 0.04 | 0.32 |
| Math Expressions relative to SFAW for teachers with: | | |
| Average View Toward Constructivist Instruction | 0.11* | 0.00 |
| View One Standard Deviation Above the Mean | 0.11* | 0.01 |
| Saxon relative to SFAW for teachers with: | | |
| Average View Toward Constructivist Instruction | 0.10* | 0.01 |
| View One Standard Deviation Above the Mean | 0.07 | 0.13 |

*Statistically significant at the 5 percent level.

Table 4
Curriculum effects by teacher explicit instruction views

| Curriculum effects | Effect size | *p*-Value |
|---|---|---|
| Investigations relative to Math Expressions for teachers with: | | |
| Average View Toward Explicit Instruction | -0.06 | 0.13 |
| View One Standard Deviation Above the Mean | 0.04 | 0.41 |
| Investigations relative to Saxon for teachers with: | | |
| Average View Toward Explicit Instruction | -0.04 | 0.32 |
| View One Standard Deviation Above the Mean | 0.05 | 0.36 |
| Investigations relative to SFAW for teachers with: | | |
| Average View Toward Explicit Instruction | 0.06 | 0.09 |
| View One Standard Deviation Above the Mean | 0.14* | 0.01 |
| Math Expressions relative to Saxon for teachers with: | | |
| Average View Toward Explicit Instruction | 0.02 | 0.59 |
| View One Standard Deviation Above the Mean | 0.01 | 0.89 |
| Math Expressions relative to SFAW for teachers with: | | |
| Average View Toward Explicit Instruction | 0.12* | 0.00 |
| View One Standard Deviation Above the Mean | 0.10* | 0.02 |
| Saxon relative to SFAW for teachers with: | | |
| Average View Toward Explicit Instruction | 0.10* | 0.01 |
| View One Standard Deviation Above the Mean | 0.09* | 0.04 |

*Statistically significant at the 5 percent level.

Table 5
Curriculum effects by classroom variation in student achievement

| Curriculum effects | Effect size | *p*-Value |
|---|---|---|
| Investigations relative to Math Expressions for teachers with: | | |
| Average Classroom Variation in Baseline Student Achievement | -0.09* | 0.01 |
| Classroom Variation One Standard Deviation Above the Mean | -0.05 | 0.31 |
| Investigations relative to Saxon for teachers with: | | |
| Average Classroom Variation in Baseline Student Achievement | -0.08* | 0.04 |
| Classroom Variation One Standard Deviation Above the Mean | -0.09* | 0.04 |
| Investigations relative to SFAW for teachers with: | | |
| Average Classroom Variation in Baseline Student Achievement | 0.03 | 0.40 |
| Classroom Variation One Standard Deviation Above the Mean | 0.06 | 0.17 |
| Math Expressions relative to Saxon for teachers with: | | |
| Average Classroom Variation in Baseline Student Achievement | 0.02 | 0.62 |
| Classroom Variation One Standard Deviation Above the Mean | -0.05 | 0.31 |
| Math Expressions relative to SFAW for teachers with: | | |
| Average Classroom Variation in Baseline Student Achievement | 0.12* | 0.00 |
| Classroom Variation One Standard Deviation Above the Mean | 0.11* | 0.02 |
| Saxon relative to SFAW for teachers with: | | |
| Average Classroom Variation in Baseline Student Achievement | 0.11* | 0.00 |
| Classroom Variation One Standard Deviation Above the Mean | 0.16* | 0.00 |

*Statistically significant at the 5 percent level.