

Curriculum Research: What We Know and Where We Need to Go

By Dr. David Steiner

In Winter 2017, the Johns Hopkins Institute for Education Policy and Johns Hopkins Center for Research and Reform in Education conducted a research review on the effects of curricular choices in K–12 education for the Knowledge Matters Campaign, a project of StandardsWork, Inc. That review,¹ available upon request at standardswork.org, surfaced several important findings, including the following:

- » Curriculum is a critical factor in student academic success.
- » Comprehensive, content-rich curriculum is a common feature of academically high-performing countries.
- » The cumulative impact of high-quality curriculum can be significant and matters most to achievement in the upper grades where typical year-on-year learning gains are far lower than in previous grades.
- » Because the preponderance of instructional materials is self-selected by individual teachers, most students are taught through idiosyncratic curricula that are not defined by school districts or states.
- » Research comparing one curriculum to another is very rare and, therefore, not usually actionable.

The overarching conclusions from the Johns Hopkins' review are that curriculum is deeply important, that a teacher's or district's choice of curriculum can substantially impact student learning, and that—as a result—the paucity of evidence upon which sound instructional, purchasing, and policy decisions can be made is a matter of deep concern and urgent need.

Following a brief review of the findings from the Johns Hopkins study, this paper examines several of the problems that persistently dog curriculum study. Given the focus throughout the Every Student Succeeds Act (ESSA) on evidence-based practice and the irrefutable evidence that curriculum matters, what must state education agencies, school districts, researchers, and funders consider when making smart, evidence-based curricular choices? We lay these problems out as follows:

- » All kinds of instructional materials are being labeled “curriculum.” Research has tended to focus on textbooks, leaving unstudied, potentially strong curricula that are not textbook based. Do we need a tighter definition of curriculum or, rather, multiple layers for a more capacious definition?
- » Because the origin and selection of curriculum is so varied, how do we quantify any differences between the effectiveness of “homegrown” versus “published” curriculum or between that which is state or district endorsed versus teacher selected?
- » Because no “taxonomy” exists of curricular features, research has not explored the elements of curriculum that really matter in student learning. We know very little about what makes a curriculum effective.

- » Distinguishing the impact of skills-building and knowledge-building features of curricula could be a particularly fruitful area of study.
- » No industry or research standards exist around fidelity of implementation. Thus, study authors often field questions about the delta between intended and taught curriculum.
- » The use of different assessments to measure progress across different schools or districts in a study often taints results from curricular interventions or renders them less than definitive.
- » Most high-quality studies of curriculum compare the curricular “intervention” with “business as usual,” but what constitutes business as usual (for example, a high-quality curriculum or nothing at all) can vary.

KEY FINDING:

Rigorous research confirms that curricular choices matter

Multiple research studies meeting the highest bar for methodological rigor find substantial learning impacts from the adoption of specific curricula. The impact on student learning can be profound.

Curriculum is a critical factor in student academic success. The What Works Clearinghouse (WWC), which is managed by the US Department of Education’s Institute of Education Sciences (IES), has identified several curricula that produce major positive effects on students’ reading. Open Court Reading, for instance, brought gains of 10 percentile points (Borman, Dowling, and Schneck, 2008), and Success for All saw 19 percentile-point gains (Smith et al., 1993). In math and science, the WWC found that the University of Chicago School Mathematics Project (UCSMP) curriculum yielded gains of up to 23 percentile points (Hirschhorn, 1993) and the TEEMSS Science gains were 24 percentile points (Zucker et al., 2008). Put differently, schools that switched from business as usual to one of these instructional methods could move students’ performance from the 50th to the 60th or even 70th percentile. When extrapolated across an entire class, grade, or school, such impacts could prove transformative.

The shift from a weak curriculum to a strong one can make an especially strong difference. Tom Kane recently reported,

Two textbooks were statistically significantly related to students’ performance—one positively and one negatively. The average student using GO Math! (Houghton Mifflin Harcourt) as their primary textbook scored 0.1 standard deviations higher (4 percentile points) than similar students using other textbooks or no textbook at all. In contrast, the average student using another textbook scored 0.15 standard deviations lower (6 percentile points) on the new math assessments. (We are not releasing the name of the second textbook because we could not confirm which edition teachers were using.) Both estimates are sizable, implying that textbook choice is a high-stakes decision (Kane et al., 2016).

The “spread” between the two textbook impacts is a 0.25 standard deviation—a 10 percentile point gain.²

The cumulative impact of a high-quality curriculum is significant. Most research studies focus on the impact of a curriculum over one or two years. But over time, even a small annual effect size of +0.10, beginning in first grade, could become an effect size of +0.60 by the end of fifth grade—approximately the equivalent of a student scoring in the 74th percentile versus the 50th percentile. A case in point is longitudinal research that tracks the performance of students receiving instruction from the UCSMP curricula. Students who were taught using this curriculum for four consecutive years (grades 7–10) outpaced comparison students by a margin of 38 percentile points—an effect size of roughly +1.16, which amounts to a stunning four additional years of learning (Hill et al., 2008).

But although the cumulative impact on student learning over several years is perhaps the greatest determinant of a curriculum's impact, most studies review academic progress over merely one academic year—very rarely over longer periods. We can see the difficulty. It may take years for instructors to master the shift from one curricular approach to another; the shift to a Common Core–based curriculum provides an obvious example. In the long run, however, the consistent use of the new curriculum over multiple years of a student's education could have a major cumulative impact. The policy implications of a state or district mandating curriculum, and therefore reaping the benefits of multi-year use of a curriculum, are significant and deserve attention.

State-based curriculum studies

The states of Florida and Indiana have tracked the results of textbook adoptions since 2009 (Polikoff and Koedel, 2017). Studies in each state found that high-quality textbooks have a positive impact on student achievement. The first study compared the effects of three different math textbooks on student outcomes in matched schools in Indiana; the second examined how well specific math subtopics were taught within a particular curriculum, compared to the same subtopics in other curricula found in matched schools in Florida.

The Indiana study examined the effect on state assessment scores of three elementary-school math programs: Saxon Math, Silver Burdett Ginn (SBG) Mathematics, and Scott Foresman-Addison Wesley (SFAW). Although SFAW produced a positive effect above Saxon Math, “the largest differences in student outcomes were found between . . . Saxon and SBG. SBG produced effect sizes of around 0.13 standard deviations (5 percentile points) of the Indiana Statewide Testing for Educational Progress exam when compared to Saxon [Math]. Effect sizes of 0.10 translate into three additional months of learning on nationally normed tests.”

The Florida study looked at elementary-school math subdomain test scores and compared the results of students who had been taught using Harcourt Math (the most common curriculum in Florida) to students who had been taught using other curricula (SFAW, SRA/McGraw Hill, Cambium, Scott Foresman Investigations, MacMillan/McGraw Hill, and Houghton Mifflin). The authors found that in data analysis and geometry, Harcourt produced statistically significant gains over the other curricula; the estimated

effect size for data analysis was between 0.092 (4 percentile points) and 0.115 (5 percentile points), and the estimated effect size for geometry was 0.108 (4 percentile points) and 0.126 (5 percentile points) for third-grade students who were taught using Harcourt in first, second, and third grades (Bjorklund-Young, 2016).

A report out of California, released in January 2017, provides still more evidence that curriculum matters. Researchers Morgan Polikoff and Corey Koedel used schools' textbook selections³ and school-, district-, and student-level data (2008–13) to compare the effects of four commonly used elementary-school math curricula: enVisionMATH California, published by Pearson Scott Foresman; California Math, published by Houghton Mifflin; California Mathematics: Concepts, Skills, and Problem Solving, published by McGraw-Hill; and California HSP Math, published by Houghton Mifflin Harcourt. They found that students who had been taught using Houghton Mifflin's California Math consistently outperformed students who had been taught using a composite of the other three on state assessments: “The effects persist across four years post-adoption and range from approximately 0.05 to 0.10 standard deviations (i.e., up to 4 percentile points) of student achievement” (Polikoff and Koedel, 2017).

KEY FINDING:

Changes in curriculum are relatively cost-neutral interventions

Because some type of curriculum is necessary to provide instruction to students, a high-quality curriculum offers a limited-cost proposition for schools. As highlighted by Whitehurst (2009),

Curriculum effects are large compared to most popular policy levers. Further, in many cases they are a free good. That is, there are minimal differences between the costs of purchase and implementation of more vs. less effective curricula. In contrast, the other policy levers reviewed here range from very to extremely expensive and often carry with them significant political challenges, e.g., union opposition to merit pay for teachers. This is not to say that curriculum reforms should be pursued instead of efforts to create more choice and competition through charters, or to reconstitute the teacher workforce towards higher levels of effectiveness, or to establish high quality, intensive, and targeted preschool programs, all of which have evidence of effectiveness. It is to say that leaving curriculum reform off the table or giving it a very small place makes no sense (Whitehurst, 2009).

The cost of placing strong curricula in the classrooms is not necessarily higher than using weak ones. As Polikoff and Koedel put it, “Textbooks are relatively inexpensive and tend to be similarly priced. The implication is that the marginal cost of choosing a more effective textbook over a less effective alternative is essentially zero” (Polikoff and Koedel, 2017).

This hypothesis corresponds to findings in two different studies.

- » The investigation of Indiana math textbooks cited above calculated the per-student cost difference between the most effective curriculum and two less-effective ones as only \$2.26 (Bhatt and Koedel, 2012).
- » Boser et al.’s 2015 report on curriculum reform identified the prices of nineteen states’ adopted elementary-school math textbooks, generated a per-student cost, and matched the cost of four specific textbooks whose effectiveness had been evaluated by the IES. They found that the difference between the highest- and lowest-quality curricula in the IES reports was only \$13 per student ((Boser, Chingos, & Straus, 2015).

The report from Boser et al. notes that, “The average cost-effectiveness ratio of switching curriculum was almost forty times that of class-size reduction in a well-known randomized experiment.”

What about the cost-benefit of using Open Educational Resources (OER) rather than published textbooks?⁴ EngageNY, for instance, is widely used and available free of charge (New York State Education Department, n/d). In 2015, Duval County began to use EngageNY districtwide. Their internal audit indicates that the district saved more than \$10 million over three years by using OER and printing the materials rather than using published curricula (Steiner, 2016).

ISSUE 1:

Curriculum can refer to many different types of instructional materials

What is a curriculum? Oxford dictionary defines it as “the subjects comprising a course of study in a school or college.”⁵ But this is not the common use in K–12 schools, where the term usually refers to the substance of what is taught and how instruction is delivered. At the capacious end of the spectrum, one finds this from the *Glossary of Education Reform*:

...Curriculum typically refers to the knowledge and skills students are expected to learn, which includes the learning standards or learning objectives they are expected to meet; the units and lessons that teachers teach; the assignments

*and projects given to students; the books, materials, videos, presentations, and readings used in a course; and the tests, assessments, and other methods used to evaluate student learning.*⁶

At the restrained end of the spectrum, curriculum refers strictly to the set of formal materials the teacher is given by the state, district, and/or school to deliver to students. The content under this definition of curriculum is derived from “published” materials such as textbooks or modules on an OER website such as EngageNY.⁷ The difficulty here is that the proportions of such materials may vary from a small to a large part of what the teacher uses. In a recent audit of the taught curriculum in a midsized mid-Atlantic district, for instance, the research team determined that the district-created “published” materials constituted less than 50 percent of the teachers’ taught curriculum. How should those materials be treated?

In the last decade, ever fewer states are prepared to define, still less require, the use of only one set of materials.⁸ This is why many states distinguish “curriculum frameworks” and “standards” from curriculum *per se*. The former lay out the skills and the level of those skills that students are expected to master, but the state can avoid being seen as overly prescriptive of the actual content that is taught in the classroom.⁹

A further definition of curriculum would have it refer to everything that a teacher actually teaches. This, however, is impossibly vague. Even in the rare case in which a teacher adheres to a scripted curriculum (as was intended by the designers of direct-instruction curriculum such as DISTAR), there are always spontaneous interactions with students that skew results positively or negatively. A final definition encompasses both the materials presented in the classroom and the teachers’ related practices, but here researchers disagree about the conditions under which such practices should be “counted.” For instance, should practices be counted only when they are described within the materials or should practices added by the teacher be included?

As an added difficulty, none of the above understandings of curriculum adequately grapple with the quality of the content that is delivered. A helpful consequence of the standards movement has been a new focus on the variation that exists in alignment between instructional materials and the state’s chosen standards. The academic rigor of curricula, and the degree to which they appropriately challenge students who are both above and below grade level, are just not addressed.

What is needed is agreement among researchers on a taxonomy that considers the multiple forms in which teachers interact with material. This taxonomy would include measures of the fidelity of implementation (in other words, whether teachers use the material in the way the author or publisher intended it; see below); the percentage of materials that lie outside of the “published” or “formal” curriculum provided by the state, district, or school; the degree to which teaching methods are prescribed within the formal curriculum (with, again, a measure of the fidelity of implementation to those methods if prescribed); the place of professional development within the curriculum; an evaluation of the academic rigor of the materials (from whatever source) that teachers deliver; and whether a given curriculum is delivered as part of a whole-school reform model, such as “Success for All,” which details multiple changes in school practices in addition to specifying a curriculum. The result of a coherent taxonomy would be a more accurate, comparable description of the curricular variables and of the outcomes of interest.

ISSUE 2:

We don’t know much about the impact of who develops and selects curriculum

The instructional landscape in American classrooms is changing rapidly. Twenty years ago, between 70 percent and 98 percent of teachers used published textbooks at least weekly (Chingos, M. & Whitehurst, 2012). Not so today. What

types of resources find their way into contemporary classrooms? A recent RAND Corporation report, *Implementation of K–12 State Standards for Mathematics and English Language Arts and Literacy*, explored this question (Opfer, Kaufman, and Thompson, 2016). Teachers reported using a variety of instructional materials from a wide array of sources: formal, published curricula and informal, online lessons; self-developed and district selected; and aligned to standards (EngageNY) and not (leveled readers).¹⁰

The preponderance of teacher-chosen and teacher-developed materials illustrates the difficulty of measuring impact, as well as the apparent rarity of sequenced study in America’s classrooms. In mathematics, for instance, teachers reported very high use of self-developed and self-selected materials in both mathematics (82 percent of elementary- and 91 percent of secondary-school teachers reporting using their own materials “at least once a week”) and ELA (89 percent of elementary- and 85 percent of secondary-school teachers reporting using their own materials “at least once a week”).

A research project led by Harvard’s Thomas Kane corroborates the high number of reported uses of teacher-created and teacher-chosen materials: 80 percent of ELA teachers and 72 percent of mathematics teachers from a representative sample in five states reported using materials developed by themselves and staff at their schools on a weekly basis and other materials less frequently (Kane et al., 2016).

Teachers also draw upon an eclectic blend of online resources, with Google.com, Pinterest.com, and Teacherspayteachers.com leading the RAND list. Their reported use of formal curricula produced by publishing companies, such as Reading A–Z and Journeys, was moderate in elementary school and rare in secondary school.

The rapid growth of online, personalized learning platforms will likely change classroom instruction further. As of yet, there exists no high-quality research on the impact of such platforms, nor do we know whether their use impacts the effectiveness of an otherwise strong curriculum. Both are important questions.

The proliferation of instructional materials presents two related but distinct issues. The first question is whether materials (of whatever kind) that have been chosen by a teacher, as opposed to having been dictated by public authority, are stronger or weaker in their aggregate impact on student learning. The second question is whether teacher-generated materials, as opposed to formal materials that represent either a published curriculum (delivered online or via traditional textbooks) or one developed by a district or school, are stronger or weaker in their respective impact on student learning. Current research does not allow us to answer either of these questions. In the first case, we simply have no research. In the second case, the enormous variety of materials created by teachers makes it impossible for research to generalize across that domain. It is certainly possible that any individual teacher or teachers could develop a curriculum that proves more effective than its published or district-created peers. However, it is perhaps fair to speculate that this is unlikely, in comparison to the most effective published materials that are the result of careful alignment with strong academic standards and thoughtful sequencing and that have been developed and vetted by master teachers and coaches.

Given that novice teachers are considerably less effective, on average, than their more seasoned peers, common sense would suggest that asking them to construct their own curriculum in addition to honing the craft of teaching will only exacerbate their challenges. Certainly, distinguishing the impact of structured curriculum versus “business as usual” on student achievement in classes with experienced and inexperienced teachers is an area in which additional research could prove deeply rewarding.

ISSUE 3:

Curriculum research hasn't examined what makes a curriculum effective

To date, research on the curriculum effect has told us little about what makes a particular curriculum or genre of curriculum especially effective or not. We encounter only occasional, anecdotal observations on this in the research. For example, in one study on an algebra curriculum (Swafford and Kepner, 1980) the reader is told that “the data suggest that the development of algebra out of real-world problems, rather than the application of algebra to real-world problems, makes working these problems more enjoyable.” The reader cannot ascertain from the research report, however, how important this “real-world-origin” quality is to the impact of UCSMP Algebra I or how influential the real-world-origin quality is compared to other features such as “readability,” which the research also identified. In another case, this time regarding the Core Knowledge curriculum, the researchers note, “The most plausible explanation for the positive effects associated with Core Knowledge is the greater curricular coherence it creates within individual schools. . . . What appears to have mattered most was the fact that the curriculum was *specified* and less so that it was Core Knowledge Content” (Stringfield et al., 2000)2000. The grounds for this judgment, however, are multiple and not cross-weighted for impact.

What is needed, where possible, is a synergetic approach to assessing effectiveness that includes the following features:

- » Both the curriculum under review and the materials used in the control group(s) should be specified. Research should evaluate the relative impact of the two across several years, following a single cohort, if at all feasible.
- » The curriculum should be classified as to its major design features and the relative weight of these different features: direct instruction, open-ended approaches, guided discovery learning, flipped classroom, blended learning, personalized learning, small-group collaborations, student-centered learning, inclusive approach, inclusion of assessments and their type, and so forth.
- » The context of the classroom and the degree of implementation should be clearly and fully described. We need to know, for instance, the school type (charter, magnet, urban, rural, selective); all available demographic data; results from a normed classroom-observation tool; additional changes in the school’s approach to teaching and learning—including professional development—that were introduced concurrently with the new curriculum; and the level of support from the school’s leadership for the researched curriculum.
- » The researchers should provide a clear evaluation of the level of fidelity of implementation (see discussion below for recommended elements). Research findings should clearly indicate the level of correlation between fidelity of implementation and the size of the curriculum effect.
- » Researchers should report the degree to which a curriculum showed success in a variety of school structures (one might term this a “robustness indicator”) and across the different student subgroups that the ESSA legislation specifies (“target population indicator”).
- » Research studies should include surveys of teachers and students to ascertain which features they found to be most effective. These findings should be cross-tabulated to the impact on student learning using nationally normed assessments.
- » Research reports should include a fiscal analysis of the material, professional development, and ancillary costs of introducing the curriculum.
- » Researchers should state which features of the curriculum, in their judgment, determine its relative impact.

ISSUE 4:

Differences in content-rich and skills-based curricula should be examined

Studies of educationally top-performing countries across the globe indicate that one of the very few characteristics they share is a high-quality, content-rich curriculum. The most extensive study, performed by a research team at Common Core, Inc.¹¹, found that a comprehensive, content-rich curriculum was the salient feature in nine of the world's highest-performing school systems as measured by the Programme for International Student Assessment (PISA). Despite the vast cultural, demographic, political, and geographic diversity of Finland, Hong Kong, South Korea, Canada, Japan, New Zealand, Australia, the Netherlands, and Switzerland, their educational systems all shared an emphasis on content-rich curriculum and commensurate standards and assessments (Common Core, 2009).

In another example, the content of the provincial curriculum of Alberta, Canada, had diminished such that by the 1970s, high schools required only two subjects for graduation—social studies and English (Heyking, 2006). In the 1990s, the government changed course (as a result of the concerted efforts of parents and civic organizations) and established curricular frameworks, created authorized resource lists for each course, and set proficiency standards that were modeled on the PISA exam (Campbell, 2004; McEwen, 1995).¹² Alberta is now among the world's most equitable and high-performing school systems (Secretary-General of the OECD, 2014).¹³ The negative evidence is also important: in his most recent book, *Knowledge Matters*, E.D. Hirsch traces the rapid decline of academic results from all sectors of French school children in the years after that country abandoned its national, content-rich curriculum (Hirsch, 2016).

Because most state standards, including the Common Core, and most state assessments, including PARCC and SmarterBalanced, are largely skills focused, many curricular materials in the United States, especially in ELA, focus on skills rather than on knowledge. This is unsurprising, given the fact that it has been notoriously difficult to agree upon which key texts students should read or which areas of knowledge they should master, particularly in middle and high school.

As a result, however, very few curricular packages are explicitly designed to be “content rich,”—that is, emphasizing a specific body of knowledge that must be mastered in addition to skills—and what few studies exist of such curricula suffer from all of the challenges discussed above. In the case of the three most rigorous studies of Core Knowledge, for instance, we find different results, although two of the three are very positive:

- » In the first study, Core Knowledge was found to have no overall impact on fourth-grade student achievement in a high-poverty urban school district but a strong, positive effect with English language learners in math and reading (effect size of +0.14 or 6 percentile points) (Datnow et al., 2003).
- » In the second study, Core Knowledge produced an average effect size of +0.17 (7 percentile points) on third-, fourth-, and fifth-grade students' performance on the Iowa Test of Basic Skills (ITBS) (Taylor, 2000).
- » In the third study, which included twelve schools in seven states,¹⁴ Core Knowledge produced an average effect size of +0.52 (20 percentile points) on the CTBS's assessments of language arts, science, and social studies but had no effect in reading or math. This study also identified an implementation effect: schools that had implemented at moderate or high levels improved their performance more than schools that evidenced low levels of implementation (Stringfield et al., 2000).

Given the international results, the study of content-rich curriculum is a compelling target for priority research.

ISSUE 5:

Questions around fidelity of implementation plague curriculum study

We have known for some time that fidelity of implementation is vital to realizing the potential of curricular impact. As one research team concluded, “[The existing] research¹⁵ ha[s] concluded that level of implementation is a significant predictor of student achievement gain” (Stringfield et al., 2000).

If we are going to rely on research to help us distinguish between the impact of strong and weak curricular models, we have to know that the research actually assesses the use of the curriculum in the classroom. But what exactly constitutes use? Researchers sometimes confirm that the curriculum they are evaluating has been implemented in the classroom with “fidelity”—meaning that teachers seem to be using the curriculum as intended. But to what degree? The answer in each case proves extremely important.

Researchers try to measure fidelity of implementation by conducting focus groups, undertaking classroom observations, and leading individual interviews with teachers. Through such measures, researchers hope to understand what content is taught, what embedded assessments are used, and whether the recommended teacher strategies are deployed. Unfortunately, however, there is no research standard for ascribing high, medium, or low levels of fidelity, still less an even more granular metric. Even research that addresses implementation can be vague. For instance, in one important study of the impact of the Core Knowledge curriculum, the researchers write, “The evaluation determined conditions under which Core Knowledge would achieve reasonably full implementation and the effects of Core Knowledge implementation in various contexts” (Stringfield et al., 2000). The question, of course, is the following: what constitutes “reasonably full implementation”? This issue is important, as in the above study, the magnitude of the curriculum’s positive impact was tied to the level or “fidelity” of implementation: “When schools implemented the sequence with greater reliability and consistency, students achieved improved outcomes on all tests. . . . Strong correlations between level of implementation and norm-referenced math and reading gains indicate that when schools implemented the Core Knowledge Sequence with greater consistency, students achieved improved outcomes in both subjects” (Stringfield et al., 2000). The research team outlined the components that make for successful implementation as follows:

Multilevel support for change was required for Core Knowledge to be successfully implemented. Successful implementation relied on instructional leadership from the principal, teacher willingness to change, and support from the district or at least a commitment that the district would enable rather than hinder long-term implementation. Successful implementation also relied upon fiscal resources to provide for the purchase of Core materials and paid teacher-planning time, as well as the organization of time, space, and professional development designed to support Core. Core Knowledge implementation was hindered by the absence of multilevel support and the presence of strong pressures to comply with state standards and accountability systems that were ill-aligned with Core Knowledge. . . . Core Knowledge was not effectively implemented in highly troubled school sites experiencing serious difficulties with school climate and discipline (Stringfield et al., 2000).

ISSUES 6 and 7:

Varying measures of effectiveness and varying business-as-usual practices make comparisons problematic

In the section on curriculum impact above, we cite research studies that compare the impact of one specific curriculum against another. Such studies are rare, particularly when it comes to randomized or quasi-experimental research, the two most rigorous research methods. Far more often, researchers compare the impact of a named curriculum with what is termed “business as usual.” Business-as-usual refers to whatever curricular materials and teaching methods happen to be used in the classrooms the researchers use as the “control” or comparison group of students. The resulting problem is clear: imagine you’re researching the impact of a particular curriculum against a control group that happens to be using another strong, evidence-based curriculum. It may be the case that even if the curriculum under review is also strong, the researchers will find little or no difference in student outcomes between the two strong curricula. However, if the research team happened to measure the same strong curriculum against a control group that happened to be using a weaker curriculum or a compilation of weaker curricula, then the same curriculum under review would register as much stronger.

A second difficulty is that schools often use multiple assessments to evaluate student learning and different schools in different states and/or districts use different assessments, or groups of assessments, than do other schools. Researchers may report that a curriculum had a positive impact in a particular school on a particular test, but not on a different test, and they may report that schools using assessment A showed a significant gain using curriculum X, but that schools in another district using assessment B showed no gains from the same curriculum. It is then left up to readers to determine the relative weight attributed to any given assessment.

When one combines these two issues—the comparing of a given curriculum to business as usual and the use of multiple assessments to judge the impact of a curriculum—the result is that almost anything can be said of any curriculum. The UCSMP Algebra I (see sidebar) offers a telling example. Three studies of UCSMP Algebra I found divergent results from the same curriculum: negative and statistically significant, negative but not statistically significant, and positive but not statistically significant.

UCSMP Algebra I: Same Curriculum, Different Results

- » A randomized study (1980) compared the performance of seven hundred students in classes that used UCSMP Algebra I with the performance of six hundred students in classes that did not, on the following assessments: the ETS Algebra I Test, a “project developed First-Year Algebra test,” an “Opinion Survey,” and a Consumer Test.” UCSMP students performed significantly below the level of the control group (effect size of -0.15 or 6 percentile points) (Slavin, Lake, and Groff, 2008; Swafford and Kepner, 1980).
- » A quasi-experimental study (1989) compared the performance of 230 students in classes that used UCSMP with two hundred students in classes that did not on the HSST Algebra assessment. After one year of use, students performed below the level of the control group, although not at a statistically significant level (effect size of -0.19 or 8 percentile points) (Mathison et al., 1989; Slavin, Lake, and Groff, 2008).
- » A quasi-experimental study (2006) synthesis compared performance of one hundred eighth- and ninth-grade students in classes that used UCSMP with a matched control group that did not, using the HSST Algebra assessment. After one year of use, UCSMP students outperformed the control group, although not at a statistically significant level (effect size of $+0.22$ or 9 percentile points) (Thompson et al., 2006).

What are we to make of these three research studies? The first study, using the most rigorous research design (randomized study), finds a negative impact of using the UCSMP curriculum on an ETS test and a variety of results on the other three tests, some of which correlated with the results on the ETS tests. But we don't know which curriculum or teaching method was used in the control groups, and we are in no position to judge the relative importance of the findings between the four assessments used. The researchers also report that “the study did not provide a realistic assessment of the ease or difficulty of implementing these materials,” thus raising the issue of fidelity of implementation discussed above.

We fare no better with the second and third studies, both of which used a quasi-experimental design. One found a negative but statistically insignificant difference for students using the UCSMP curriculum from the control group – but on a completely different assessment (the HSST algebra test). The other used the same assessment as the second study (the HSST algebra test) and, this time, found a positive but statistically insignificant difference. We do not know, in either the second or the third study, which curricula the two control groups of students were using. It is simply not possible to draw secure conclusions about the UCSMP Algebra I curriculum from this body of research.

Another interesting point on the subject of effectiveness measures: the impact of curriculum has been almost exclusively evaluated based on academic test scores. Rare exceptions are curricula from such programs as Advanced Placement exams, which are designed to substitute for first-year college courses. Under ESSA, states are required to quantify nonacademic measures of student learning, which could include such traits as perseverance or creativity. Currently, we have no research that focuses on the impact that specific curricular materials and accompanying teaching methods might have on the positive acquisition of such skills. As such, this is another area ripe for study.

A NEW AGE FOR CURRICULUM

Over the last twenty-five years, education has seen multiple waves of large-scale federal, state, and local reform initiatives that aim to improve student performance, especially the performance of disadvantaged students. These have included everything from reducing class size and boosting digital learning to creating common standards and have encompassed such movements as choice, accountability, and teacher quality. The sad reality is that although these reforms have been expensive, expansive, and often divisive, the performance of twelfth graders on the 2015 National Assessment of Educational Progress (the “nation’s report card”) was mediocre—unchanged in math since 1992 and actually lower in reading. Stagnate student performance that leaves some 60 percent of high-school graduates unprepared for postsecondary training or schooling, the persistence of racial and socioeconomic achievement gaps, and the seeming futility of reform efforts all suggest that American public education is badly in need of a new path forward.

In this deeply challenging environment, strong research should make curriculum a top priority. What we teach isn't some sidebar issue in American education: it *is* American education. The track record of top-performing countries, early evidence of positive effects from the faithful implementation of high-quality curricula here in the United States, and the persistent evidence that our classrooms are underchallenging our students at every level compel us to put the materials that we use to teach at the core of serious education reform.

DR. DAVID STEINER *is the Executive Director at the Johns Hopkins Institute for Education Policy.*

Endnotes

1 The literature review includes research compiled from widely used and respected sources such as Johns Hopkins University’s Best Evidence Encyclopedia, the U.S. Department of Education’s What Works Clearinghouse, and the Visible Learning Project, led by education scholar John Hattie. References to international evidence is included where appropriate.

2 See Marzano Research Laboratory’s (2009) “What is an Effect Size?” (Marzano Research Laboratories, n/d).

- 3 All schools in California must report their textbook selections to the state.
- 4 The Hewlett Foundation’s explanation of OER: “OER are teaching, learning, and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use and repurposing by others. Open educational resources include full courses, course materials, modules, textbooks, streaming videos, tests, software, and any other tools, materials, or techniques used to support access to knowledge” (Hewlett Foundation, n.d.).
- 5 <https://en.oxforddictionaries.com/definition/curriculum>
- 6 <http://edglossary.org/curriculum/>
- 7 <https://www.engageny.org>
- 8 See, for example, *Education Week’s* 2015 analysis of the decline in the number of states that have formal textbook-adoption policies (Gewertz, 2015) or, as an example, Indiana’s policy brief (2012) on its deference to local textbook-adoption policies: <http://www.doe.in.gov/sites/default/files/curriculum/textbook-faq-6-21-2012.pdf>.
- 9 For one example, see Massachusetts, which sets out extensive curriculum frameworks that nevertheless stop short of assigning specific content or materials (Massachusetts Department of Elementary and Secondary Education, n.d.).
- 10 Leveled reading attempts to match students’ level of reading skill and frustration with an “appropriate” text. The Common Core State Standards support the use of challenging texts for all students, a position that opponents dismiss (Pondiscio, 2014). The RAND reports consider leveled readers to indicate lack of standards-aligned materials and practices.
- 11 The nonprofit organization Common Core, Inc. (not associated with the Common Core State Standards) changed its name in 2014 to Great Minds.
- 12 Alberta also redistributed property-tax levies to equalize funding across the province and expanded school choice to include Jewish, Evangelical, Inuit, and home schooling to the existing pluralistic system.
- 13 Currently, Alberta Education is revising its curriculum and is, it seems, drawing more heavily upon skills than knowledge: <https://education.alberta.ca/curriculum-development/how-is-curriculum-developed/everyone/curriculum-development-process/>.
- 14 Colorado, Florida, Ohio, Maryland, Tennessee, Texas, and Washington.
- 15 “From the Sustaining Effects Study (Berman and McLaughlin, 1977; 1979), Disseminating Efforts Supporting School Improvement (DESSI) (Crandall et al., 1982), and Special Strategies (Stringfield et al., 1998)” (Stringfield et al., 2000)2000.

Citations

- Bhatt, R., & Koedel, C. (2012). Large-Scale Evaluations of Curricular Effectiveness: The Case of Elementary Mathematics in Indiana. *Educational Evaluation and Policy Analysis*, 34(4), 391–412.
- Bjorklund-Young, A. (2016, April 11). High-Quality Curricula: A Cost-Effective Way to Boost Student Learning. Retrieved from http://education.jhu.edu/edpolicy/commentary/hq_curricula
- Borman, G. D., Dowling, N. M., & Schneck, C. (2008). A multisite cluster randomized field trial of Open Court Reading. *Educational Evaluation and Policy Analysis*, 30(4), 389–407.
- Boser, U., Chingos, M., & Straus, C. (2015). *“The Hidden Value of Curriculum Reform.”* Washington, D.C.: Center for American Progress. Retrieved from <https://cdn.americanprogress.org/wp-content/uploads/2015/10/06111518/CurriculumMatters-report.pdf>.
- Campbell, D. E. (2004). The Civic Implications of Canada’s Education System. In P. Wolf, S. Macedo, D. J. Ferrero, & C. Venegone (Eds.), *Educating Citizens*: (pp. 186–220). Washington, DC: Brookings Institute.
- Chingos, M., & Whitehurst, R. (2012). *Choosing Blindly: Instructional Materials, Teacher Effectiveness, and the Common Core* (Brown Center on Education Policy). Washington, D.C.: Brookings Institution.
- Common Core. (2009). *Why we’re behind: What top nations teach their student but we don’t*. Washington, D.C.: Common Core.
- Datnow, A., Borman, G. D., Stringfield, S., Overman, L. T., & Castellano, M. (2003). Comprehensive School Reform in Culturally and Linguistically Diverse Contexts: Implementation and Outcomes from a Four-Year Study. *Educational Evaluation and Policy Analysis*, 25(2), 143–170. <https://doi.org/10.3102/01623737025002143>
- Gewertz, C. (2015, February 18). States Ceding Power Over Classroom Materials - Education Week. *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2015/02/18/states-ceding-power-over-classroom-materials.html>
- Hewlett Foundation. (n/d). Open Educational Resources. Retrieved January 21, 2017, from <http://www.hewlett.org/strategy/open-educational-resources/>
- Heyking, A. J. V. (2006). *Creating Citizens: History And Identity in Alberta’s Schools, 1905-1980*. University of Calgary Press.
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research. *CDEP Child Development Perspectives*, 2(3), 172–177.
- Hirsch, E. D. (2016). *Why Knowledge Matters: Rescuing Our Children from Failed Educational Theories*. Cambridge, MA: Harvard Education Press. Retrieved from <http://www.barnesandnoble.com/w/why-knowledge-matters-e-d-hirsch-jr/1123826690?ean=9781612509525>
- Hirschhorn, D. B. (1993). A longitudinal study of students completing four years of UCSMP mathematics. *Journal for Research in Mathematics Education*, 24(2), 136–158.

- Kane, T., Owens, A., Marinell, W., Thal, D., & Staiger, D. (2016). *Teaching Higher: Educators' Perspectives on Common Core Implementation* (p. 56). Cambridge, MA: Center for Education Policy Research, Harvard University. Retrieved from <http://cepr.harvard.edu/files/cepr/files/teaching-higher-report.pdf?m=1454988762>
- Marzano Research Laboratories. (n/d). Appendix B: What is an Effect Size? Retrieved from https://soltreemrsl3.s3-us-west-2.amazonaws.com/marzanoresearch.com/media/documents/pdf/AppendixB_DTLGO.pdf
- Massachusetts Department of Elementary and Secondary Education. (n.d.). Massachusetts Curriculum Frameworks: Current. Retrieved from <http://www.doe.mass.edu/frameworks/current.html>
- Mathison, S., Hedges, L. V., Stodolsky, S., Flores, P., & Sarther, C. (1989). *Teaching and learning algebra: An evaluation of UCSMP Algebra*. Chicago: University of Chicago School Mathematics Project.
- McEwen, N. (1995). Educational Accountability in Alberta. *Canadian Journal of Education / Revue Canadienne de L'éducation*, 20(1), 27–44.
- New York State Education Department. (n/d). EngageNY: Frequently Asked Questions. Retrieved January 21, 2017, from <https://www.engageny.org/frequently-asked-questions>
- Opfer, V. D., Kaufman, J., & Thompson, L. (2016). *Implementation of K–12 State Standards for Mathematics and English Language Arts and Literacy: Findings from the American Teacher Panel*. Santa Monica, CA: The Rand Corporation. Retrieved from http://www.rand.org/content/dam/rand/pubs/research_reports/RR1500/RR1529/RAND_RR1529.pdf
- Polikoff, M., & Koedel, C. (2017). *Big bang for just a few bucks: The impact of math textbooks in California* | *Brookings Institution* (Evidence Speaks No. #5). Washington, DC: Brookings Institution. Retrieved from <https://www.brookings.edu/research/big-bang-for-just-a-few-bucks-the-impact-of-math-textbooks-in-california/>
- Pondiscio, R. (2014, September 24). Leveled reading: the making of a literacy myth. Retrieved from <https://edexcellence.net/articles/leveled-reading-the-making-of-a-literacy-myth>
- Secretary-General of the OECD. (2014). *PISA 2012 Results in Focus: What 15-year-olds know and what they can do* (p. 44). Paris: Organisation for Economic Co-Operation and Development. Retrieved from <http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf>
- Slavin, R. E., Lake, C., & Groff, C. (2008). *Effective programs in middle and high school mathematics: A best evidence synthesis*. Baltimore, MD: Johns Hopkins University, Center for Data-Driven Reform in Education.
- Smith, L. J., Ross, S. M., Faulks, A., Casey, J., Shapiro, M., & Johnson, B. (1993). *Wayne, IN SFA results 1991-1992*. Memphis, TN: Center for Research in Education Policy.
- Steiner, D. (2016, October 27). Duval and High-Quality Curriculum: A Conversation with Superintendent Nikolai Vitti. Retrieved from <http://edpolicy.education.jhu.edu/wordpress/?p=940>
- Stringfield, S., Datnow, A., Borman, G., & Rachuba, L. T. (2000). *National Evaluation of Core Knowledge Sequence Implementation: Final Report*. Baltimore, MD: Johns Hopkins University Center for Research on the Education of Students Placed at Risk. Retrieved from <https://eric.ed.gov/?id=ED451282>
- Swafford, J. O., & Kepner, H. S. (1980). The evaluation of an application-oriented first-year algebra program. *Journal for Research in Mathematics Education*, 11, 190–201.
- Taylor, J. G. (2000). *Core Knowledge: Its impact on the curricular and instructional practices of teachers and on student learning in an urban school district*. Nova Southeastern University, Fort Lauderdale, FL.
- Thompson, D. R., Senk, S. L., Witnosky, D., Usiskin, Z., & Kaeley, G. (2006). *An evaluation of the second edition of UCSMP Algebra*. Chicago, IL: University of Chicago School Mathematics Project.
- Whitehurst, G. (2009). *Don't Forget Curriculum* (Brown Center Letters on Education) (p. 12). Brookings.
- Zucker, A. A., Tinker, R., Staudt, C., Mansfield, A., & Metcalf, S. (2008). Learning science in grades 3–8 using probeware and computers: Findings from the TEEMSS II project. *Journal of Science Education and Technology*, 17(1), 42–48.