



The Science of Standards-Based Education

A standards-based model of reform has dominated public education for 30 years. Under the Every Student Succeeds Act (ESSA), it will continue to dominate education policy. But is that model working? State boards of education share an intrinsic interest this question. While

there are many ways to investigate it, one approach that shows promise treats standards-based education as a testable theory and offers empirical tools that show policymakers, administrators, and teachers how well it is functioning in their state, district, school, or classroom.

New tools let educators compare actual classroom instruction against standards and point them toward how to align the two.

John Smithson

Box 1. Measuring the Efficacy of Standards-Based Education

So is the hypothesis testable? Consider the requirements. If you were researching the question, you would want a pretty detailed, comprehensive description of the content that students had had the opportunity to learn. You would also want a detailed, comprehensive description of the content being assessed (knowledge and skills).

Inevitably, there would be ambiguities to live with, like not being able to identify every opportunity that every student in the sample group had to learn some knowledge or skill. Likewise, you would have to deal with an incomplete description of the assessment, assuming both prerequisite and ancillary knowledge and skills embedded in it have been neither fully described nor documented.

Once collected, these descriptions of instruction and assessments would have to be compared with student performance to determine whether students' opportunities to learn had helped to predict or explain their assessment scores. Keep in mind this would all have to be done on a relatively large scale, across many classrooms and preferably many different tests. Of course, you would also have to find some way to connect the description of a particular student's opportunity to learn with that student's score on the assessment, or better yet, the student's scores on each item. This is no small challenge, nor is the outcome as certain as it may have first seemed—not because the theory isn't viable but because of the measurement challenges involved.

■

While prior achievement and poverty remain the strongest, most consistent predictors of student achievement, close behind is an additional predictor related to something that teachers affect.

Since *A Nation at Risk* was published in 1983, state and federal education policymakers have been using standards to drive more rigorous expectations for student outcomes. Beginning with graduation requirements for increased course taking in mathematics and science, the standards movement quickly moved to “national” content standards offered by one or another professional association, such as the National Council of Teachers of Mathematics, which released its standards in 1989. During the 1990s, state after state adopted academic content standards for most if not all academic subjects. No Child Left Behind (NCLB) further intensified these reforms and expanded the federal role in solidifying them, and ESSA retains the essential structure of standards-based reform and its attendant policy levers, even as it modifies the federal law based on the lessons learned from NCLB.

Though typically viewed as a policy tool, standards-based reform also represents a theory of education, possibly a testable one (see box 1). Does a standards-based approach to education really work? It appears to be self-evident that students are more likely to learn what they are taught than what they are not taught. But then

it also appears that the sun travels around the earth, the surface of which appears relatively flat.

A 1998 study by Gamoran and Porter et al. (including myself) demonstrated an important connection in education that may seem obvious.¹ Because we focused primarily on investigating the impact of alternate strategies for increasing student success in high school mathematics, we were able to demonstrate a predictive link between opportunity to learn and content coverage (regardless of pedagogical practice). While that was not the major finding or focus of the study, it was important nonetheless. Not because of what it said about education (yes, students can learn what they are taught) but what it said about our instrumentation. And those findings have been replicated twice—on large-scale populations, across academic subjects, and across states—using high-stakes standardized assessments.² The results point to a variable that, even after controlling for prior achievement and the effects of poverty, consistently predicts student achievement in large-scale applications.

Consider the implications. Our findings indicate that while prior achievement and poverty remain the strongest, most consistent predictors

of student achievement, close behind is an additional predictor related to something that teachers affect. That additional predictor is students' opportunity to learn, defined as the content of instruction that students actually experience—what we call the enacted curriculum.

For states that were paying attention, these studies indicated that efforts to better align curriculum to academic standards could be expected to lead to higher achievement scores. But how would an educator or state board member know that instructional alignment has improved or that efforts to make it so have been successful?

Enter the Surveys of Enacted Curriculum (SEC), which marries the Porter content instrument and a traditional-style survey to comprehensively collect teachers' opinions and beliefs, professional development experiences, instructional practices, and content coverage. The content portion of the survey uses a three-step process to collect information from teachers regarding content coverage. The first step is to identify those topics from the SEC content topic list that the teacher covered to some extent during the reporting period. In the second step, teachers report how much time they spent on those topics for which they reported coverage. The response options are "slight coverage (less than a class period/lesson)"; "moderate coverage; (1 to 5 classes/lessons)"; and "sustained coverage (more than 5 classes/lessons)". In the final step, teachers report on the relative emphasis they give to five categories of expectations for student performance for each of the topics they covered. The labels and examples of the five categories vary slightly by subject area, but all reference essentially the same general categories of performance expectations: 1) recall; 2) perform procedures; 3) demonstrate understanding; 4) analyze information; and 5) integrate/synthesize information. It is this connection between topic and cognitive demand that sets SEC data apart and provides its predictive power.

SEC data can be used to create a detailed description of instructional content that can then be compared with a similar description of the state's learning standards or relevant assessment in order to observe similarities and differences between the descriptions. Figure 1

provides an example of two content maps, in this case displaying descriptions of the academic content embedded in the Common Core State Standards for mathematics (on the left) and language arts and literacy (on the right). Similarities and differences between the content descriptions of standards maps like these and the content descriptions of instruction provided anonymously by teachers can be captured quantitatively, and at a very detailed level, to paint a picture of the extent to which classroom instruction and the relevant content standards overlap.

The idea is to provide teachers, administrators, and policymakers information on instruction and its relation to standards and performance expectations in order to make informed decisions about how to increase student achievement. Given the link we have established between opportunity to learn and achievement gains, it seems reasonable to expect that efforts to bring standards and instruction into sync will lead to improved student learning as measured by state assessments (assuming the standards-based system is functioning as expected).

Michigan's Statewide Approach

No state has done more to align curriculum using these tools than Michigan. Although Michigan was an early adopter of the SEC, like many states it has used the tools episodically as new initiatives arise. So I was pleasantly surprised when Karen Ruple, program manager for the MI Excel Statewide System of Support at the Michigan Department of Education, called a few years back expressing interest in using SEC data with Michigan's designated priority and focus schools. The Excel program provides Michigan's most challenged schools support in improving student achievement through an array of services, including the SEC. Sustained by the efforts of a well-trained cadre of curriculum specialists and others drawn from the state's intermediate support agencies, Michigan today uses SEC extensively to improve curricular alignment under its MI Excel umbrella.

Realizing that the SEC represented a fairly complex data set that required substantial training and support before it could be used to effectively engage teachers, Michigan invested in the development of web-based training modules

How would an educator or state board member know that instructional alignment has improved or that efforts to make it so have been successful?

Figure 1a. Common Core Mathematics Standards, All Grades

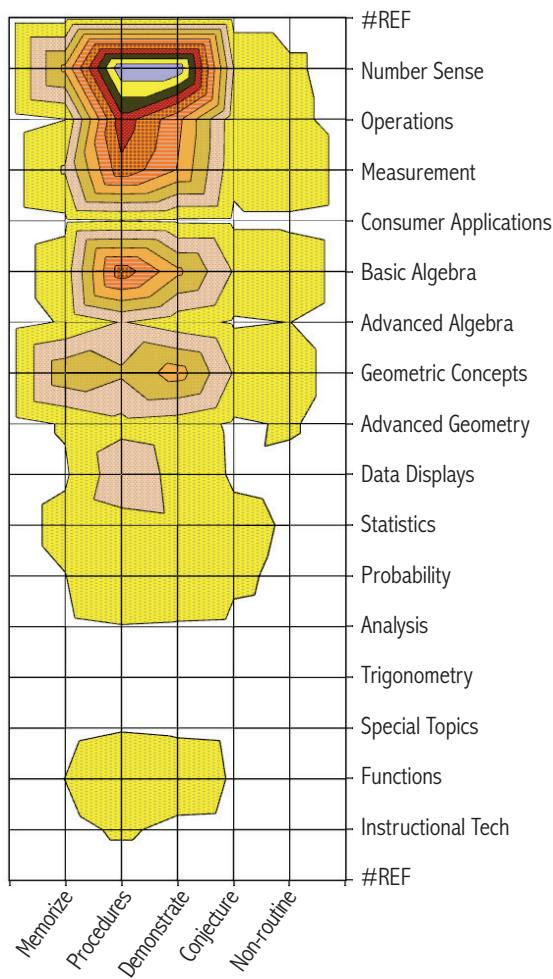
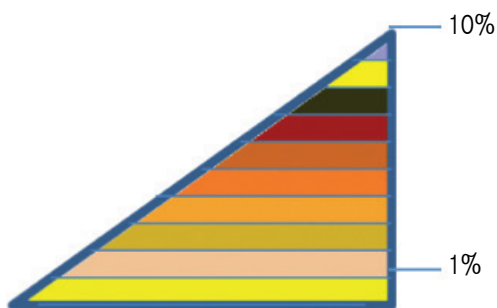
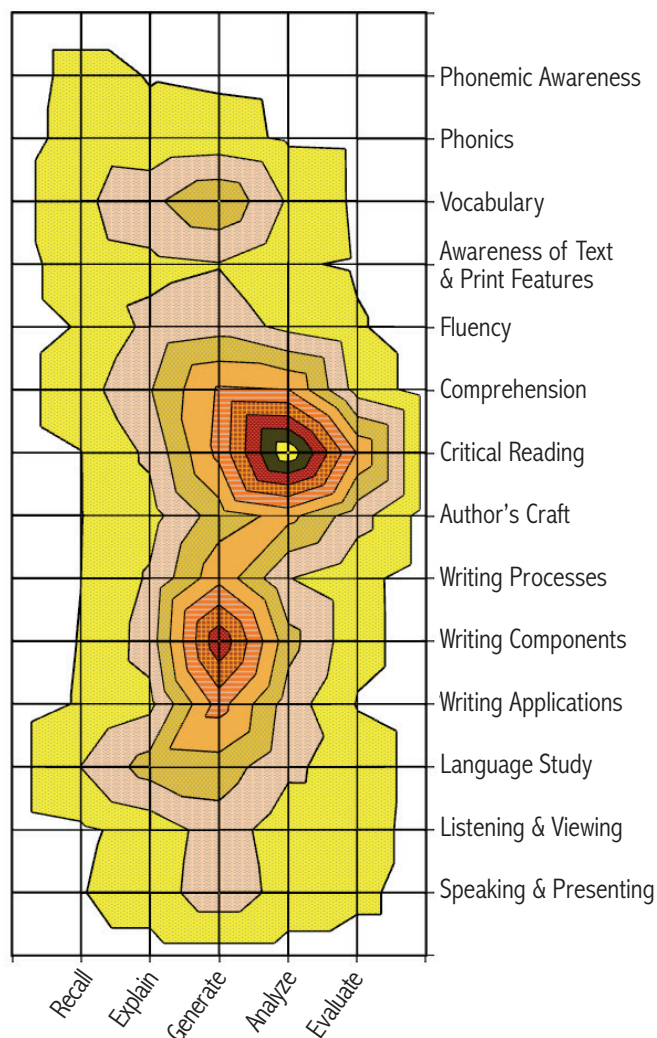


Figure 1b. Common Core State Language Arts, Reading & Literacy Standards, All Grades



The vertical or y-axis for each of these charts provides a list of topic areas based on the SEC content taxonomy for each subject. The horizontal or x-axis reports a series of cognitive demand categories. Measurement is made at the intersection of these two dimensions, and results are reported as the relative emphasis of academic content across the full standards document. The z-axis reports the relative emphasis for each intersection of topic and cognitive demand, so that the sum across all intersections equals one for each chart. The most emphasized content for mathematics across grades K-12 is Number Sense at the cognitive demand levels of procedural and conceptual (demonstrate) understanding. The most emphasized content for language arts, reading, and literacy across grades K-12 is Critical Reading at the cognitive demand level of analyze information.

for teachers and administrators. The department also pulled together a leadership team from the state's intermediate service agencies to provide additional support to teachers and administrators in getting the most out of the SEC information.

The impact of these efforts on instruction and students' opportunity to learn can be tracked through the results reported by the SEC system each spring, and changes in instructional alignment can be examined over time.

A couple of caveats: First, even clear evidence of increased curricular alignment might be discounted simply because it is based in part on teacher reports. Without added evidence that the increased alignment was also associated with increased student performance, the data's validity can be questioned. Second, even with evidence of a link between instructional alignment and achievement, it is important to distinguish the alignment target. If the predictive validity we see is based on instructional alignment to the *assessment*, we have only established the validity of the survey and assessment instruments and fallen short of testing our bigger initial question about the viability of the standards-based education *system*. True, a link between instructional alignment to a test and achievement does provide evidence to support the fundamental assertion of education that students are more likely to learn what they are taught (all else being equal), but standards-based education goes one step further, and it's a big step.

To date, there is anecdotal evidence of school-based increases in instructional alignment, combined with increased achievement scores, leading in some cases to schools' removal from the list of Michigan's most challenged schools.³ However, the state has not yet undertaken a systematic study of the link between instructional alignment and student achievement. For an example of a state looking closely at the link between instruction, assessment, and student achievement, we have to turn to North Carolina.

North Carolina Examines Assessment Validity

North Carolina, like Michigan, was one of the original states to develop and use the SEC

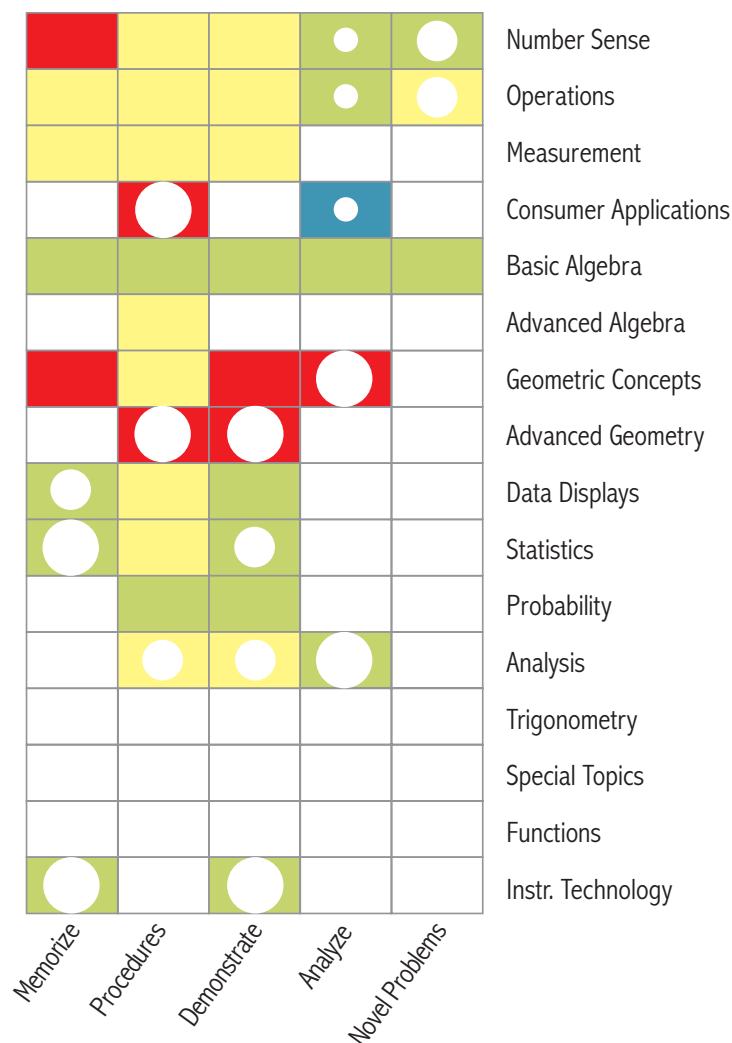
instruments and has participated in several multistate studies using SEC data, most recently a study into the learning opportunities of English language learners.⁴ Not long after the completion of that study, Tammy Howard, director of accountability services for North Carolina's Department of Public Instruction, contacted me about the possibility of conducting a validity study of the state's assessment program. In 2015, the department approved a three-phase empirical study of whether North Carolina's assessment program functions as intended (i.e., increased alignment to standards will lead to higher achievement on the assessment).

My colleagues and I are two-thirds done. In Phase I, we established the alignment characteristics of all of the assessment forms for all of the assessments for all of the assessed grades in math, reading, and science by conducting content analyses of each test. In Phase II, we surveyed a representative sample of teachers and provided each participating teacher a confidential, personalized alignment report comparing their reported instruction to the state's standards for their own reflection and in consideration of their participation in the study, along with a series of webinars for district and school leaders on the purpose of the study and the data requested from teacher surveys.⁵ In Phase III, under way now, the state is providing the necessary student assessment data (with full confidentiality protocols in place) to conduct the tests for predictive validity.

Once we receive the student performance data, we will conduct two types of statistical tests. The first is to see if alignment of instruction to the assessed content helps to explain test performance, and the second is whether instructional alignment to state standards predicts test performance.

Failure on the first test would lead us to question the validity of the instructional and assessment descriptions (but not the original hypothesis of whether students are more likely to learn what they are taught, which in addition to being self-evident has already been confirmed and replicated, as noted above). If we get a positive result on the first test but fail to predict achievement with the second, then we have a quite

Figure 2. SEC Skills Analysis Chart



The skills analysis chart above reports student performance results for assessed content (in this case a series of high school mathematics classroom assessments). Red indicates that the majority of students failed the items related to that content, while green indicates the majority of students responded correctly to items assessing that content. Blue indicates that more than 85% of students responded correctly to the assessed content. Yellow indicates that about as many students answered items related to that content correctly as incorrectly. White tiles indicate content that was not assessed. Circles inside a tile indicate that results are based on a limited number of items. The larger the circle, the less reliable the results reported for that content.

different situation. Under those circumstances, the instrumentation will have been validated, but the system would have failed. There could be any number of reasons why this might be so, and a negative finding here would beg further investigation to determine the nature of the problem. Even then, we would have gathered sufficient

descriptive data to be reasonably confident that a careful review would likely reveal areas where one or another component of the study did not function as expected, thereby suggesting changes in method or instrumentation likely to increase our understanding of the connection between instruction and achievement.

Even if the North Carolina test is successful on both fronts, the question will remain whether other state education agencies will be emboldened to pursue this type of systemic evaluation themselves. The potential payoff is huge, but the risk to those responsible for accountability and assessment program design or implementation is not trivial. States willing to undergo the rigors of such an evaluation, however, will have demonstrated a commitment to assessing the accountability of the system itself, and not just a part of it. I suspect such willingness will go a long way toward garnering support for—and faith in—the state’s accountability and assessment programs among local superintendents, administrators, and teachers.

What’s Ahead

Call me an optimist, but I am most excited about what it means to succeed in the effort to validate North Carolina’s assessment program because that would not be the end of the story but just the beginning.

Consider the possibility of all teachers having available to them training on opportunity to learn and data on instructional alignment such as that available to teachers in Michigan and with the kind of achievement data soon to be available in North Carolina. Performance data collected on an assessment instrument that has been analyzed using SEC methodology can be reported in a unique data display that clearly highlights distinct areas of academic content where students are succeeding or not. Such information, in conjunction with the opportunity-to-learn data, provides teachers a powerful tool for targeting instruction, helping them determine what works, what doesn’t, and what to do about it (figure 2).

The utility of such data extends further, though. Many factors affect student achievement. While SEC does not directly address many aspects of students’ learning opportunities, it can help point the way toward a better understanding of what actually is going on in the classroom.

Qualitative research, while critical in increasing our understanding of classroom dynamics and student learning, is rarely able to demonstrate a solid connection to achievement gains

on any large-scale basis. While much of the difficulty lies in the challenges of conducting large-scale qualitative studies, another important element is that the effects being sought are generally buried in the statistical noise or measurement error reflected in the data set. Instructional alignment, or opportunity to learn, can assist in reducing a large amount of that noise, thereby increasing the sensitivity of the qualitative measures of interest in demonstrating an effect on achievement.

Much of this is hypothetical and perhaps overly hopeful. But we have done the necessary groundwork and have a pretty good sense of the predictive capacity of the SEC measures. I therefore eagerly await the opportunity to put all the pieces together from the work in North Carolina and Michigan to demonstrate not only the viability of the standards-based approach to education but also the potential the SEC model offers in support of a science-based approach to education. ■

¹A. Gamoran et al., “Upgrading High School Mathematics Instruction: Improving Learning Opportunities for Low-Achieving, Low-Income Youth,” *Educational Evaluation and Policy Analysis* 19, no. 4 (1997, Winter).

²John Smithson and Cristina Collares, “Alignment as a Predictor of Student Achievement Gains,” paper presented at the 2007 Annual Meeting of the American Education Research Association, Chicago; John Smithson, “Measuring Opportunity to Learn: Benefits for Research and Practice,” paper presented at the 2013 Annual Meeting of the American Education Research Association, San Francisco.

³There have even been cases where schools chose to continue using the SEC tools even after removal from the list despite having to foot the bill (as the state covers SEC use only among listed schools).

⁴John Smithson, “Report on Cross-State Analyses of Students with Disabilities’ Opportunity to Learn Standards-Based Content and Its Impact on Student Achievement,” report to the Council of Chief State School Officers as part of US ED grant #S368A100013 (Washington, DC: Council of Chief State School Officers, 2013).

⁵Note that North Carolina does not employ a curricular alignment initiative comparable to Michigan’s. The purpose of the SEC data collection effort in North Carolina was to collect a representative sample of instruction across the state for use in the study.

Consider the possibility of all teachers having available to them training on opportunity to learn and data on instructional alignment.

John Smithson is project director of the Surveys of Enacted Curriculum Project at the Wisconsin Center for Education Research in Madison, Wisconsin.