

What We Test Matters

Let's not shortchange deeper learning

Stuart Kahl | October 2017

Notwithstanding controversy over federal involvement, statewide assessment systems provide accurate information for schools and districts to use to improve their instructional programs. It's important to acknowledge what statewide tests are—and are not. They are general achievement measures that sample from a whole year's worth of content within a subject and grade, so state tests are generally not designed to be particularly diagnostic at the individual student level. They don't provide teachers with information they should act on the day after testing. However, they do provide good information on the effectiveness of school programs—data on local student performance relative to that of students in like communities and statewide, as well as test results for subgroups of students or on subdomains of a subject area. Those results can be used to identify students or content areas that need additional attention in instruction.

Impact of State Tests

While these tests shouldn't be the sole source of evidence for teacher evaluations or student advancement, they do have stakes associated with them. Schools and school personnel are judged, at least in part, based on their students' performance. And research has shown that the content and nature of state tests can significantly affect the enacted curriculum and instruction. This finding has been particularly evident in the case of high-stakes tests that rely heavily or even exclusively on multiple-choice items, which tend to shortchange higher-order thinking skills and deeper learning. Logically, teachers emphasize in their instruction the kind of knowledge covered by the state tests and emulate those measures in their own testing.

How Can We Test Deeper Learning?

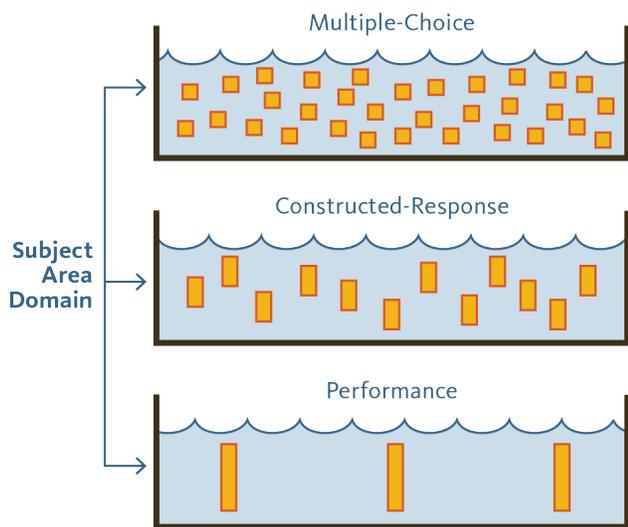
Despite the emphasis on deeper learning in college and career readiness curriculum standards that states have adopted, state assessments still do not do justice to these standards. Studies of the alignment between state tests and the curriculum standards on which they are based typically show

“ . . . a state assessment program should include . . . some extended performance tasks that are part of instructional units . . . ”

that the tests inadequately assess higher-level cognitive processes.

While rich performance tasks could address that problem, the time and costs required are a limiting factor in the implementation of performance assessments as part of the end-of-year “external” tests. Performance components included in these tightly controlled, on-demand, efficient tests are often nothing more than a few thematically related constructed-response questions in math or a single writing task for English language arts. Such isolated tasks are not enough to (1) weigh meaningfully in the assessment results, (2) contribute significantly to the reliability of the larger assessment, or (3) adequately capture the spirit of the curriculum standards in terms of deeper learning.

If we want teachers to give adequate attention to deeper learning—i.e., students' ability to apply foundational knowledge or basic skills—then we should include measures of deeper learning in our state assessments. In my mind, that means a state assessment program should include a component made up of some extended performance tasks that are part of instructional units or projects implemented during the course of the school year. With teacher scoring, these would yield results of immediate use to teachers. Tried-and-true tasks from the state and a scoring audit procedure could make these results suitable to contribute to state accountability assessment scores along with the data from a more traditional, efficient (and shortened) end-of-year test. For program evaluation, the latter test would still be necessary. Results should generalize to the larger domain of a subject area at a grade, from which representative items and tasks are sampled. The graphic below illustrates how different item/task types cover a domain.



An Ideal Balance

My ideal state assessment program includes all three item or task types. While I've already mentioned the negative impacts a high-stakes, all-multiple-choice test can have, including some multiple-choice items makes sense. Each of these items addresses a very small portion of the domain in isolation, but collectively they can provide broad coverage of the domain. Constructed-response questions—

“ BINGO, we have a highly reliable accountability assessment addressing both deeper learning and foundational knowledge and skills.”

particularly those that are worth up to four points and take a student 8 to 10 minutes and a fair amount of space to answer—cover material that several multiple-choice items could cover, but require higher cognitive processes. A few constructed-response questions can be administered in a relatively efficient, end-of-year test along with multiple-choice questions.

Curriculum-embedded performance assessments, part and parcel of instructional units, are the answer to including performance tasks in statewide assessments. Within units that span several days or even weeks, several activities could lead to student work that can be scored for summative purposes. Based on research and years of experience, I believe three such units and associated performance tasks would be ideal. Looking back at the graphic, one can see that by themselves, three performance tasks would not provide adequate coverage of a subject domain; but with an end-of-year component that has a couple of constructed-response questions and 30 or so multiple-choice items . . . BINGO, we have a highly reliable accountability assessment addressing both deeper learning and foundational knowledge and skills. But wait, there's more! This design provides many efficiencies of time and cost because of the curriculum-embedded component, teacher scoring, and shortened end-of-year test; AND it has positive instructional impact.

Most importantly, this two-component assessment system can cover the “right stuff”: It can measure a representative sampling of the full range of content and cognitive processes called for by college and career readiness standards.