

[Education Week's blogs](#) > [Assessing the Assessments](#)

Honestly Prioritizing Assessment Design for Instruction

By Invited Contributor Listed Below on April 4, 2014 12:01 AM | [No comments](#)

Today's guest contributor is **Neal Kingston**, Professor, Psychology and Research in Education; Director, Achievement and Assessment Institute, University of Kansas

"I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail." Abraham Maslow, 1966.

In educational assessment we are surrounded by experts and practitioners wielding hammers. Ask the experts to create an assessment used for accountability purposes and they will take out their hammer and bang out a test that is highly standardized. Students will all answer the same questions (mostly multiple-choice) on the same days (typically in one week at the end of the school year). Test specifications will be tightly controlled so that everyone is tested with the same amount of each type of content. Much care will be spent on statistical issues like item calibration, test equating, reliability, and comparability. These tools, once thought important to help us reach our goal of better student learning, have become goals in their own right.

Ask a testing company to create an assessment used to inform instructional decisions and they will take out the same hammer and create a test that looks and feels pretty much like the accountability test. It will be developed the same way using the same statistical analyses. They may use advanced statistical methods like item response theory or adaptive testing, but to what end? Perhaps the tests will be a bit shorter or cover a narrower range of content, but there is no evidence that they will support student learning any better than the accountability test. The largest difference is likely to be an increased number of unreliable subscores for which there is no evidence of utility. And why should we be surprised? - They were created using the same hammer.

Meanwhile, we have provided tens of thousands of teachers with hammers and shown them how to hit nails. And perhaps we have lied by telling them that screws are the same as nails and encouraging them to hammer the screws down too... but perhaps I have taken this metaphor far enough.

No one test can serve multiple purposes equally well. We need to choose the most important goal for a testing program and focus on achieving it. Secondary considerations should get secondary priorities. We need to identify the evidence we need to support the inferences we must make and support the actions we best take. Rather than develop tests that first and foremost support accountability decisions and then as an afterthought try to figure out how to extract a few morsels of instructionally useful information we should develop assessments that are designed to help teachers make good instructional decisions and figure out how those tests can be used to reasonably support accountability decisions.

High on the list of lessons learned from No Child Left Behind is that when a test is used in a high-stakes accountability system teachers will teach to the test, so it behooves us to have a test worth teaching to. This means designing the test from the ground up based on how students learn (including especially how different students learn differently), modeling good instructional activities, and providing information that supports decision-making during instruction. Once we start there, and without allowing ourselves to compromise those desiderata, we can develop and apply appropriate psychometric models to address the secondary goals.

Teachers have a huge amount of work to do. Not all teachers can be equally expert at all aspects of teaching, including the elicitation of feedback they need to optimally instruct their students. Most teachers will benefit from tools designed to support them, especially a structured assessment tool based on sound cognitive science. These tools should support teacher decision making but not attempt to replace teachers as decision makers. The ways in which we have chosen to standardize the testing experience is part of the problem, but there have long been better approaches to similar problems in other fields - we are just not developing or using them in educational assessment. The concept of mass customization - the well-structured designing of flexibility into a product to allow it to meet individual needs - needs to be applied to educational assessment. Cognitively Based Assessment **of, for, and as** Learning (**CBAL**) is one research initiative trying to apply some of these principles but CBAL is an initiative not a product. In Fall 2014 the Dynamic Learning Maps Alternate Assessment (**DLM**), developed at the University of Kansas, will be the first operational assessment system to embody these principles to serve the learning needs of individual students. With these projects as models, perhaps it is time to put down our hammers and to start focusing on teaching and learning again.

Neal Kingston

Achievement and Assessment Institute

University of Kansas

Categories: [Validity](#) [High-stakes testing](#) [School accountability](#) [International assessments](#) [Test use](#)

Tags: [achievement growth](#) [achievement status](#) [standardized achievement tests](#) [standardized metrics](#)