



Stuart Kahl, Ph.D.
Founding Principal
Measured Progress, Inc.

Technology and the Future of Assessment: Pitfalls and Potential

In looking to the future of educational assessment through a technology lens, it is easy to envision some pretty impressive applications of technology being used to help gauge what students know and can do. Such things as gaming and simulations create all kinds of possibilities. However, it would do shapers of the future no harm to keep recent history in mind and perhaps think in terms of both the near future and the more distant future.

Part 1: The Glass Is Half Empty

Fifteen years ago, Black and Wiliam told the world of the remarkable academic gains that can be accomplished by the effective use of the multi-step instructional process called “formative assessment.” Unfortunately, for several reasons, the term came to mean “frequent testing” to many. Because timing of the evidence-gathering step (during instruction) and immediacy of feedback are important to the process, online delivery of multiple-choice tests is what many chose to do in the name of “formative assessment.” The research on feedback indicating that rich, descriptive feedback was the most effective for formative purposes was largely ignored, as were other steps in the larger process.

All this was going on during the early and middle years of NCLB, when the requirements of high stakes testing at many grades and of quick turnaround of results to accommodate parental choice decisions caused many states to reduce or discontinue their use of expensively scored non-multiple-choice formats, including constructed-response questions and more extended performance tasks.

Efficiencies of time and cost ruled the day. There are now students and teachers whose only educational experience has been during the NCLB era. And with increasingly higher stakes associated with state test results, it is not surprising that the tests teachers use in their classrooms emulate the state tests. Is it any wonder there are concerns now about students’ lack of higher order thinking skills and the ability to apply foundational knowledge and skills to more complex real-world problems? Students’ exposure to such problems and teachers’ exposure to (seeing) actual student work enabling more effective feedback have been diminished significantly over the past decade and a half.

“ Is it any wonder there are concerns now about students’ lack of higher order thinking skills and the ability to apply foundational knowledge and skills to more complex real-world problems? ”

Business people and policy makers are now calling for deeper learning and college and career readiness in our pre-college students. Accordingly, the Common Core State Standards (CCSS) for English Language Arts and Mathematics and the Next Generation Science Standards do the same. The two major state assessment consortia, PARCC and Smarter Balanced, federally funded through the Race to the Top Program, are charged with assessing student achievement relative to the CCSS. Early plans for the consortia's "next generation" assessments included all-online administration of tests and extensive use of both extended performance assessments and automated computer scoring of student work. However, both consortia have scaled back their plans for performance assessment because of high stakes assessment concerns of feasibility, security, and efficiency, as well as psychometric considerations. PARCC chose not to count through-course performance assessments to be administered at different points during the school year. Smarter Balanced, whose early working group on performance assessment envisioned multi-day, project-like activities, some of which could involve group work, reduced the scope of their performance component to more traditional on-demand tasks that can be administered in back-to-back periods in the same day.

Historically, automated scoring of student work has been more "trusted" in the area of writing assessment. The leadership of one consortium had hoped that by the time their assessments went operational, advancements in artificial intelligence scoring would make scoring of student higher order work more defensible in content areas other than writing. However, while existing automated scoring systems used in educational testing can sort, match, count, etc. to score proxy measures of various writing traits, they have not yet proven themselves in the other

disciplines. In fact, there have been instances in which this situation has led to the development and use of lower level questions that can be scored by the existing automated systems.

The consortia both plan on using technology-enhanced items. These are test questions that use such techniques as drag-and-drop and hot spots and other approaches for students to interact with the computer to respond to the questions. The intent is for these questions to assess skills not readily assessed via multiple-choice questions but still be scored by computer-based scoring solutions. These, of course, have a lot of promise. Yet writing TEIs that measure something more than can be measured by multiple-choice items can be challenging. Many so-called TEIs fail to accomplish that. For example, asking a student to drag a car image to a place on a hill where it has the greatest potential energy measures nothing more than an item asking students to pick the one graphic out of four depicting a car positioned on a hill where it has the greatest potential energy.

Many schools and school districts across the country are far from online-assessment-ready, at least with respect to the high-stakes accountability assessments. The consortia's decision to proceed with mixed modes (online and paper) reflects a more realistic sense of where schools are relative to their technology capacity and how long it could take them to get up to speed for the PARCC and Smarter Balanced online tests. Furthermore, the online systems themselves face challenges that go beyond the capabilities of the local schools. Security issues, limited testing windows (of time), and the need to accommodate very large numbers of simultaneous users have made the creation of problem-free test delivery systems very difficult. (Of course, the number of simultaneous users is a problem at the school level as well.)

Thus far in this paper, a fairly sobering picture seems to have been painted. Online testing programs commonly used in our schools have tended to focus on low-level, isolated knowledge and skills, and many schools are far from ready to participate in high-stakes online testing. The delivery systems for large-scale accountability assessment themselves are certainly not perfected. But let's look at the situation regarding technology and assessment from another perspective. In many ways, both the near future and the distant future look very bright.

Part 2: The Glass is Half Full

There is no question that technological advancement will continue to lead to amazing capabilities in the more distant future. Our only concern about that time should be remaining mindful that technology is a tool making our work easier and more efficient, but that it should not be a driving force determining our behavior merely because of its capabilities and limitations. We may have lost sight of this somewhat in recent years as we became more and more entrenched in assessing lower order knowledge and skills. The focus of this section, however, is on the near future of online assessment—what we can do now and what we can do in the near future with existing capabilities.

First, let's look at different levels of knowledge and skills through a well-known example: drivers' tests. In addition to meeting qualifying requirements regarding age, vision, etc., a candidate for one's first driver's license typically must pass two important tests: a "written" (actually multiple-choice) test and a behind-the-wheel driving test. To pass the first, a candidate must master the contents of the driver's manual, learning the rules of the road—the meanings of signs and signals, stopping distances, right-of-way rules, and other aspects of traffic law and safe driving. In preparing for this test, candidates do whatever they must to learn the content, including asking others to "drill" them, letting them know when they answer a question wrong.

It is doubtful that anyone would want to be on the road with drivers who got their licenses on the basis of the "written" test alone. While studying the manual, candidates with learners' permits also practice real driving, applying what they've learned from the manuals and developing other skills necessary for them to be able to do that. This is what deeper learning is all about.

The primary focus of testing, including online testing, in the NCLB era—both high stakes accountability testing and school/classroom-based testing – has been on foundation knowledge and skills, not unlike the focus of the drivers' written test. While our assessments go a little farther than that by addressing conceptual understanding and including isolated "mini-applications," they have not effectively tapped deeper learning by engaging students in the kinds of activities consistent with the broader goals of education involving more complex, real-world problem solving, for instance.

There are some promising programs emerging that involve curriculum-embedded performance assessments (CEPAs), for example: the Ohio Performance Assessment Pilot Program (OPAPP) and Quality Performance Assessment operated by the Center for Collaborative Education in Boston. CEPAs can be viewed as instructional units that include multiple learning and evidence-gathering activities, some of which may lead to products or performances that are evaluated for formative purposes and some that are scored for summative purposes. It is

“ ...in the way of technology, we have what we need right now for very effective performance assessment, both curriculum-embedded and not.”

important that assessment systems give significant attention to deeper learning at all levels from classroom assessment to statewide accountability testing. In fact, at the state level, one would think deeper learning and performance assessments would be the major focus, the latter being more direct measures of competency relative to the broader goals of education. At the local level, of course, diagnostic testing relative to foundational knowledge and skills is equally important. Online test delivery systems as they exist today can play and are playing a significant role in this diagnostic testing.

Effective performance assessment, neglected for some time, is on the rise. Fortunately, with greater awareness of the concept of alignment of tests to standards, current underused capabilities of technology, and lessons of the 1990s authentic assessment era, we know how to do it. While a more distant future for educational technology will be exciting and productive, one should not underestimate the time it takes for schools on a large scale to change or even have the ability to change. Yet in the way of technology, we have what we need right now for very effective performance assessment, both curriculum-embedded and not. In conjunction with these assessments, we should see more of the following:

- students, sometime individually and sometimes in groups, engaged in learning and assessment activities that involve digital research, exploration, collaboration, and organization tools;
- students producing scoreable products making use of on- and off-line tools to publish, storyboard, map, create videos, make movies, screen-capture, etc.; using presentation tools and apps for interactive whiteboarding, screen-casting, and multi-media presentations;
- students storing their work and submitting it for evaluation via digital portfolio; and
- teachers and others—anywhere—scoring student work and auditing scores using distributed scoring systems.

Clearly these applications of technology are far from new, but the NCLB era has seen too little use of them. While advances in technology will undoubtedly produce some exciting new tools, we need to make much greater use of some of the ones we have now and, at the same time, make sure the new tools don't drive us to practices that inhibit deeper learning. Educational reformers are calling for changes in the ways many teachers and students spend their time, and existing tools identified above can go a long way in supporting those changes.

About Measured Progress

Measured Progress is a Dover, New Hampshire-based, not-for-profit organization dedicated to improving student learning and instruction in the standards-based classroom. Since 1983, Measured Progress has successfully partnered with more than 30 states and hundreds of districts across the nation in support of assessment programs that have affected millions of students. Measured Progress develops state- and district-level assessments and assessment materials, and is the nation's leading provider of alternate assessment for students with cognitive disabilities. A not-for-profit organization, Measured Progress is deeply committed to its mission of helping educators improve student learning by connecting assessment, data, and instruction.

It's all about student learning. Period.

