

EDUCATION'S DIGITAL FUTURE

Autumn 2012

**THE NATURE OF ASSESSMENT
IN DIGITAL EDUCATION**

Engin Bumbacher, Becca Constantine, David Lopez,
Simon Wiles, Benje Williams, and Johnny Winston

INTRODUCTION

Popular demand for evidence of student learning has existed for centuries. Assessment is central to our educational culture; it is where we operationalize what we want out of education. As education engages many stakeholders, it serves a variety of goals; assessment is used as a means to determine the outcomes of education with respect to those goals.

Philip Nagy argues that assessment serves three roles (Nagy, 2000):

- Gatekeeping ("in which assessment determines who is granted a privilege such as admission or graduation")
- Accountability ("in which assessment is used to decide if schools are working well")
- Instructional Diagnosis ("in which assessment is used to find out what students do and do not know, and what to do about it")

While the role of gatekeeping is the one with the longest history (Nagy, 2000), the role of assessment as a measure for accountability of schools districts, schools, and teachers has become increasingly important in recent years. Initiatives such as the PISA studies - an international evaluation system for students' literacy in several core domains - and the No Child Left Behind Act (NCLB) in the United States have enshrined assessment for its accountability value. NCLB, for example, is based on the premise that setting high standards and establishing measurable goals can improve individual outcomes in education. To ensure positive outcomes, NCLB requires standardized testing of students' basic skills as the means of accountability of the school districts, schools, and teachers. Thus, there is increasing focus on standardized tests not only as a measure of student learning but also as a measure of accountability and evaluation.

This idea of "killing two birds with one stone" has a strong impact on the workings of the schools at the classroom level, not necessarily for the good. This not only motivates schools and teachers to "teach to the test," but also defies what many would consider the main purpose of assessment: instructional diagnosis. For example, the fact that standardized test scores correlate strongly with students' family incomes suggests that the tests do not measure student learning as they should.

The nature of classroom assessment in a given classroom, school, or district defines what is valued and what students are learning. Standardized tests, unfortunately, do not necessarily measure what we value as educators, education researchers, and activists. An almost religious belief in numbers, both as quantifications of educational processes and as objective measures of quality, has led to a distorted view that educational problems such as the achievement gap and low student achievement can be solved by standardized tests as supposedly rigorous evaluation and assessment tools.

We do not focus in this paper on assessments used to give grades or to satisfy the accountability demands of an external authority. Instead, we emphasize assessment as one important part of instruction that can support and enhance learning. As Ray Pecheone, Executive Director of the Stanford Center for Assessment, Learning, and Equity put it: "What we really want to do in accountability, in part, is build the kind of assessment that is *worth teaching to*."

Ray Pecheone articulated this argument during the Panel on the Future of Assessment held during the November 13th session of Stanford University's Education's Digital Future (EDF) class. Three experts in the field of assessment—Ray Pecheone, Ed Haertel (Professor at the School of Education), and Edys Quellmalz (Director of Technology-Enhanced Assessments & Learning Systems at WestEd)—came together with the EDF community to discuss their work on alternatives to the current assessment systems. The panel focused on the rapidly changing technologies that bear new affordances to build assessments that go beyond basic skills to capture more complex skills that are considered "21st-century competencies."

As the panelists shared, past and future technological developments offer exciting new perspectives and tools to create assessment systems that focus again on student learning instead of evaluation. However, there is no magic bullet. As Richard J. Shavelson wrote in *A Brief History of Student Learning Assessment*, "One of the most dangerous and persistent myths in American education is that the challenges of assessing student learning will be met if only the right instrument can be found—the test with psychometric properties so outstanding that we can

base high-stakes decisions on the results of performance on that measure alone” (Shavelson, 2007).

In light of these developments, we look in this paper at promising aspects of technology as a tool for assessment of complex skills, but also at limitations of digitally assessing competencies such as creativity or interpersonal skills. We quote extensively from EDF students, who responded to the assessment panel on a digital Piazza forum, to describe assessment’s current standing and some benefits and limitations of digital technology as an assessment tool.³

ASSESSMENT’S CURRENT STANDING IN TRADITIONAL & DIGITAL CONTEXTS

Assessment today, in both traditional and digital contexts, is highly focused on testing content knowledge. As Albert Lim effectively summarizes:

We assess how much students have learned, be it the sciences, languages or humanities, and whether they are able to apply their knowledge in different contexts. Content knowledge is easy to assess using pen and paper tests. Today, as soft skills or 21st century skills become more important, we need to develop ways of assessing these skills and competencies. Written tests can still work for skills in the domain of "Cognition", e.g. critical thinking, reasoning, and current tests and examinations already test for these skills. However, written tests will be inadequate in assessing skills in the domains of "Interpersonal" or "Intrapersonal" skills. For example, it is not possible to use written tests to assess skills like collaboration, responsibility or initiative.

As Albert suggests, significant changes need to be implemented in order to ensure that we are effectively teaching the right skills and competencies. However, as he later points out, what is also missing in the conversation is a more holistic discussion about the purpose of assessment:

³ The EDF students whom we quote in these pages are all graduate and undergraduate students at Stanford University who posted these reflections in direct response to the Panel on the Future of Assessment. The reflections responded to an assignment prompt developed by the EDF teaching team which asked students to consider the task of assessing “21st century competencies” such as critical thinking, metacognition, and conflict resolution.

Without sorting out the purposes of assessment, and how skills need to be assessed (as opposed to assessing content knowledge), it is difficult to plan or discuss how assessment would change digitally.

Einstein once said that "Not everything that can be counted counts, and not everything that counts can be counted." Taking a step back, are we so obsessed about measuring everything we teach, that we need to devise ways of measuring every soft skill? Yes, there are merits to knowing exactly how much a student has progressed in a certain skill, but it might not justify the efforts and time teachers will need to spend to get that precise measurement.

Albert's critical view is echoed by Simon Wiles: "What can be tested ends up being the defining characteristic of what should be taught. You end up putting the cart before the course.... We can't teach this because there's no way to test it."

Emily Schneider digs deeper into this issue:

Most traditional assessments look at the end product of a learner's experience - an essay, a creative work, a science report, a score on a standardized test. But the majority of the deep learning competencies are processes, not products. This makes perfect sense, as these competencies are meant to be transferable across domains, and be strategies that are drawn on in order to learn new things and participate in new groups. This also means that the processes need to be measured, and the final product should be treated as a validation of the success of the processes.

Hallie Fox also reinforces the problems of assessing the end product, using her own experiences as a teacher to support this concern:

The role of assessment is not only to determine what your students have learned, but assessments also determine what you will teach. As a teacher, I often started with the end goal first: the test and standards. I then planned backwards to fill in unit plans, monthly and weekly topics, and daily lessons. What to teach and how to teach it fundamentally changes when the nature and type of assessments change.

As Engin Bumbacher delicately summarizes, “Assessment is still seen and designed as the process that follows an instructional intervention, as a process that merely measures the state of the education of students without impacting them.”

Fortunately, however, as Hallie illustrates, “the move to formative and interim assessments is a positive one.” Unfortunately, this move is not without complications, as Jennifer Der Yuen points out:

Assessing soft skills is a challenge in traditional classrooms; assessing them in digital learning environments is even more challenging. From my experience teaching, I've measured soft skills based on rubrics driven by observations and self-evaluations. However, the opportunity to "observe" interaction is quite limited (different?) in a digital classroom.

In the next two sections, we consider the benefits and challenges of assessment in light of this current tests-before-goals environment.

ASSESSMENT IN THE DIGITAL ENVIRONMENT: BENEFITS & VALUE

Educational assessments can be understood as a process of documenting and validating learning within the constraints of time and academic resources. Consider this hypothetical scenario: a classroom has a one-to-one student teacher ratio; the teachers can constantly observe and assess the students' learning and provide immediate feedback. Extremely high-quality assessment would be possible within this classroom environment. This scenario is obviously unreasonable, however, due to the time and resource costs that it requires on the part of the teachers. Technology allows us to imagine a time in the near future when it will be possible to achieve such rich, guided assessments at low costs. A wonderful thing about digital technology is that it scales at extremely low costs; it might provide in-depth, personal assessment (like having a teacher by your side) without the cost of employing a large teaching staff.

Many EDF members considered such possible benefits of digital technology as they imagined how technology might help teachers assess 21st-century competencies. Rene Kizilcec details how virtual reality technology might be used to assess interpersonal conflict resolution:

Interpersonal conflict resolution is a very important social skill in life. Some research suggests that kids, in particular, forego opportunities to practice their interpersonal skills, including conflict resolution, because a considerable amount of face-to-face (F2F) communication has been replaced by computer-mediated communication (CMC). The problem with this shift in communication, is twofold: first, CMC reduces the richness of social interactions (Media Richness Theory), and second, different interactions occur in CMC than in F2F communication, specifically, we see fewer expressions of negative emotions. This brings me back to the practicing conflict resolution 'by doing'. In an environment in which nonverbal behavior does not exist beyond emoticons, and in which negative expressions of emotion are more uncommon, there is little opportunity to practice conflict resolution.

Why VR? I believe that the highly practical skill to resolve conflicts should also be assessed as such. In virtual reality you can give each learner the same conflict situation with the same agents, or vary the level of conflict difficulty. Learners have a safe environment to practice and get tested, where nobody gets emotionally hurt, while the level of immersion and emotional responses on the side of the learner are very similar to a real world situation.

Molly Bullock describes how video chats might be used to practice and assess oral communication:

Let's imagine a task where students have a "buddy" somewhere else in the country or world...Next, let's sat the student buddies video chat every 6 weeks and exchange other communications between these "face-to-face" discussions. They are given both prompts to use during discussion and a set of issues to discuss and resolve....On assessment day, this communication task is automatically recorded

by the learning software. If this takes place in the not too distant future, the videos are machine coded using some combination of the criterion above. Both now and in the future, the video can be used for local analysis as well. Students are given the opportunity to watch the video and assess their performance (maybe by tagging spots in the video or just giving scores on a rubric with explanations for their scoring). Perhaps the students score their partner or other peers in their class. Teachers can also assess performance using either tagging software or a clearly explicated rubric. Together, they set improvement goals for certain skills based on observed performance and machine code.

One common theme running through the assessment discussion was that digital technology can be leveraged to collect personalized data *during* the learning process, whereas traditional non-technological assessments (multiple-choice tests, etc.) collect data mostly *after* the learning has been done and thus deliver delayed feedback. Shuchi Grover expresses this eloquently and articulates how certain students may flourish in online assessments:

Clearly, dispositions are challenging to teach and ever more challenging to assess. However it might be easier to nurture and assess some of these dispositions in online learning environments than in the traditional classroom. I hark back to Candice Thille's lecture and her demonstration of the granularity of individual student information that is available to the teacher based on performance in the adaptive learning environment. Let's take grit as an example. Perseverance to complete a task successfully has been shown to be linked to self-efficacy (as the article also suggests). I reckon adaptive learning environments will be able to do this much better than a teacher charged with working with a bunch of kids with varying needs and ability levels. Technology-enabled personalized learning environments can select the level of difficulty of problems to target at students, help them to be successful at task completion, and gradually amp up the difficulty of problems all the while helping students believe in themselves and their abilities. Teachers can play a crucial role in face-to-face interactions by creating a classroom culture that fosters the growth mindset.

Leadership and communication are other dispositions that can be nurtured through technology-mediated learning environments - especially for some students. Online discussion forums are equalizers in a way physical classrooms are not. Those who are unable to be heard in the physical space (due to several social factors that are at play in the typical high school classroom) may find a voice in an online space - maybe even a MOOC (where the interaction is with a cohort of strangers). Online spaces for interaction could help provide some students the opportunity to take on leadership roles in a way they never could in their brick and mortar classroom. Archives of online interactions could give teachers exceptional levels of insights to measure and evaluate student communication, leadership roles, and even group collaboration and teamwork.

Success in online learning has been shown to be linked to self-regulation. Students need to take more responsibility for their own learning. Hybrid/blended learning classrooms at the high school level can be leveraged to develop and assess this soft skill as well.

Edys Quellmalz of WestEd showed some examples of the types of new assessments possible with digital technology. She began by describing some of the limits of current technology-enhanced assessments. She explained that current assessments often are under-documented, lack conceptual clarity about their target knowledge and skills, and are designed as afterthoughts to eye-catching learning environments. She focused on one project called SimScientists that provides an alternative model. SimScientists uses dynamic interactive environments to assess science process skills like inquiry and model-building. For example, students can answer questions about an ecosystem by tinkering with the populations and variables involved, testing hypotheses and developing explanations. She noted that this is a time of promising changes in the development of project-focused assessments that use multiple representations and in the assessment of "new literacies".

Indeed, technology might hold the key to creating low-cost, scalable, and rich assessment environments. Even if we can't assess all complex competencies digitally, we can still improve assessment in education overall by redistributing which work is done by people and which work is done by computers. If we can automate some tasks (like rote grading and assessment) and redistribute them to computers, then we can free up human teachers to perform tasks that computers are not yet good at, such as complex assessments of subtle competencies. Thus, an increased digital presence in education may directly and indirectly improve assessment of skills and 21st-century competencies.

ASSESSMENT IN THE DIGITAL ENVIRONMENT: RISKS & CHALLENGES

Among the posts on Piazza, a number of responses to the Panel on the Future of Assessment were keen to draw attention to issues, problems, and potential disadvantages that are associated with the move to a digital platform for educational assessment. Participants were invited to reflect especially on the challenges inherent in an attempt to assess the degree of acquisition of 21st-century competencies.

Jennifer Der Yuen reminds us that assessing such skills in a digital context requires new solutions:

Assessing soft skills is a challenge in traditional classrooms; assessing them in digital learning environments is even more challenging. From my experience teaching, I've measured soft skills based on rubrics driven by observations and self-evaluations. However, the opportunity to "observe" interaction is quite limited (different?) in a digital classroom.

Perhaps the single biggest challenge that emerges from the posts concerns the nature of the "21st century competencies" under discussion and their amenability to digital forms of assessment. Kristen Howell explains:

It is pretty straightforward to measure mastery of a skill or concept, but really assessing the deeper learning competencies related to higher order understanding

and future application is definitely a huge challenge. It is very hard to capture a student's true understanding using any format, particularly when you take into account the wide range of students in a typical public classroom in terms of culture, content knowledge and language.

Engin Bumbacher puts the question succinctly:

[T]echnological assessment approaches that are mainly evaluative and less formative might have limitations when it comes to the development and assessment intrapersonal 21st century skills. Skills like critical thinking or metacognition are generic skills, often independent of the context, and thus have to be developed across disciplines. How can technology handle this?

Drawing on his experiences at the Graduate School of Business, Tyler McNally is able to frame the kinds of questions which need to be asked to evaluate fully students' acquisition of interpersonal competencies, yet he is aware of a bridge which is yet to be crossed:

The biggest challenge is how to assess the answers to these questions [concerning the development of interpersonal skills] ... the questions are impossible to answer in multiple choice form.

Tanner Veal points to technological limitations which he feels may prove ultimately to be intractable:

Among the potential issues, perhaps the biggest is the technological one. Are the natural language and image processing algorithms sophisticated enough, and is the hardware fast enough, to provide the kind of feedback [we require]?

Many respondents expressed concern that digital modes of assessment might prove to be less suitable for certain kinds of students. David Ayrton Lopez, for example, expresses concern about the applicability of digital assessment for younger learners:

Most of my teaching experiences have occurred in primary school classrooms—a setting where assessment is difficult to implement, and perhaps even more difficult to sustain. How do you assess students at a young age, particularly given that they are unlikely to have experience with formal sit-down methods?

Disparities among students with respect to enthusiasm for and competence with digital tools and environments is at the heart of reservations articulated by Liam Aiello:

I am concerned by the fact that certain students would feel more comfortable demonstrating their creativity in another format, perhaps less comfortable with the use of digital tools. I'm also concerned with the students' unwillingness to share reflections so openly, in an online collaborative space, which would further limit this tool's usefulness for assessing metacognition.

Liam goes on to introduce a further complication, one which is perhaps inherent in the nature of assessment more broadly conceived, but which presents specific challenges in the digital context:

And lastly, as many others have pointed out, we assess for what we as teachers have prepared our students to do. Teachers must carefully consider how they are cultivating these competencies so that students' creativity isn't something that arises in the moments that students are assessed; neither would it be effective for teachers to develop a lens on such broad and nuanced competencies that defines them solely by the behavioral criteria on display during the assessment platform. Formative assessment of these competencies might be better accommodated by the powerful data-collection allowed by digital ed tools, but the complexities of defining and teaching for these competencies still persists.

Emily Schneider makes a closely related point, reminding us that the digital environment can constrain just as much as it can liberate. In her comments about assessing the skill of collaboration, she offers a key caveat:

[When] collaboration is happening only in the digital world (without participants being physically co-present), it is important to remember that the options that you give people for communicating with your software limit the types of interaction that you can measure.

Educators and other test designers must maintain constant vigilance to ensure that digital assessment techniques don't unduly constrain the responses that students can make. In addition, they must maintain an awareness of the importance of the context in which those responses are made. Tanner Vea expresses sensitivity to this role of context in the assessment of collaborative skills and competencies:

I also suspect that what constitutes “good” collaboration and communication may be very dependent on the domain—or even the specific task at hand. Finally, especially in the parent-child context, uneven knowledge may make it harder to parse out what aspects of language use and behavior are the result of poor collaboration and what are the result of differences in expertise (either with the domain or the technology itself).

Although Engin Bumbacher expresses optimism that even the “soft skills” that Jennifer Der Yuen alludes to may be amenable to algorithmic analysis, he also sees that the very complexities which make this possible may also serve to diminish the role of the teacher and the context they provide—with deleterious effects:

[P]urely technological assessment systems might be complex enough to assess certain 21st century skills (see “Technology-Based Assessment of 21st Century Skills”, <http://www.wested.org/cs/we/view/rs/1216>), but they will probably also be obscure to teachers. This is problematic if ... a student's performance can only be assessed adequately when knowing about his or her context, when knowing what the teacher has given as assignments or what projects have been done. In a purely technological approach to assessment, the teacher's role gets further reduced to the "executer" of the assessment infrastructure - i.e. he has to properly

prepare the students, and make sure they take the tests, etc.—and valuable knowledge might get lost.

Much of the discussion surrounding digital assessment highlights the advantages and potential gains that new methodologies and practices promise, even at the same time as it draws out and articulates some of the challenges, which must be faced to deliver on those gains. There is no guarantee, though, that the inevitable changes will be for the better. Galina Meyer sounds a note of caution:

However, this over-accurate data [collected in digital assessments] might profoundly affect how students approach learning, probably in a negative way. The ‘backpack of information’ becomes a very legitimized tag for students, probably affecting self-esteem and self-identity in ways we can’t anticipate. I also worry that quantifying the learning process too much shifts the importance of education to another, different game of numbers.

There are organizational and practical challenges to be overcome in realizing effective solutions. Despite being generally very optimistic about the possibilities for truly useful learning environments, Shuchi Grover is very aware that they will require the cooperation and collaboration of teams of professionals working across traditional disciplinary boundaries:

[I’m p]retty sure though that a Silicon Valley startup full of software engineers will not be able to do design such a learning environment—it’ll need the help of teachers, counselors, researchers and developmental psychologists who understand child and adolescent development to pull it off :-)

CONCLUSION

While we cannot predict the entire future of digital learning, assessment will undeniably retain a large role within that future. Regardless of how the rest of the educational landscape changes, schools will still be held accountable for student learning and some form of assessment will

remain inextricably linked to measurement of performance and standards. At the same time, our system of proficiencies and values is rapidly evolving to reflect societal changes. New 21st-century skills have come to replace antiquated models of citizenship and overall competency. This new skill set places a greater emphasis on real-world competitiveness and is designed to focus more on learning than on standardized evaluation.

Emerging educational technologies seem to show great promise in providing new ways of measuring student progress and assisting in the implementation of assessment. The allure of multiple modalities may make these technologies entirely irresistible; the corresponding flexibility they offer teachers may lead to more meaningful, customized progress reports. Educators will also have new opportunities to offer real-time support to students and to adjust their instruction accordingly. Digital platforms offer unparalleled opportunities for collaboration, which is proving to be invaluable in a globalized world, and they expand access to education and to high-quality assessment tremendously: it would be impossible to mention digital technologies without lauding their ability to deliver information to mass audiences across borders and socio-economic capacities.

Perhaps the greatest challenge for digital means of assessment is that current forms of assessment are still imperfect. Simply put, even if educators were armed with the most sophisticated tools, there is no “secret recipe” for what a good method of measuring student progress looks like. Even if there were, it would likely have a fairly short shelf life (i.e., the introduction of “new standards” such as the 21st-century skills would disrupt it quickly). Also, as many students alluded to on Piazza, acclimatization is a major concern when implementing any new technology. Education is a notoriously slow-moving field, making learning curves for adjustment to digital technology even steeper. The fact that different students have different technological capabilities muddles the future of new digital assessment methods. While some students may be very used to using computers, others may be using them for the first time. This discrepancy poses a serious threat to the integrity of the assessments.

Although the prospects may seem ambiguous, further exploration into the world of digital assessment can only bring about new knowledge. Additionally, it is possible that an

amalgamation of digital and traditional methods could bring about a highly effective way to implement new 21st-century skills in the classroom. Regardless, technology appears to be here to stay and, so far, experimentation in digital assessment has hinted at being too promising to ignore.

BIBLIOGRAPHY

- Nagy, P. 2000. "The Three Roles of Assessment: Gatekeeping, Accountability, and Instructional Diagnosis." *Canadian Journal of Education*, Vol. 25, No. 4:262-279.
- Shavelson, R.J. 2007. "A Brief History of Student Learning Assessment." Association of American Colleges and Universities.

