E-Learning and Paper Testing: Why the Gap?

Assessment has failed to keep up with advances in technology used by students and teachers alike

By **Scott L. Howell**

It is the last class of the term, and the zoology professor has scheduled this class period to review for the final exam. In the question-and-answer session, a student expresses confusion over an earlier reading in the text. The professor invites the class to don their 3D glasses as she projects from her laptop a simulated muscle contraction with an animated color overlay that traces the series of chemical-induced reactions triggering the contraction. The student, with personal digital assistant (PDA) in hand and computer in lap, manipulates his own copy of the simulation to trigger the contraction event and says he now understands. Relieved, he thanks the professor—and then asks, "So I understand it now, watching a 3D animation of the whole process, but how is a test on paper in the testing center going to be fair? It's 2D, static, no color, and my hand always cramps when I write too long because I am so used to keyboarding everything."

The zoology professor sighs, shutting off her computer as the rest of the class vocally supports the student's questions. "You're right—after a class where I use the best technology possible to help you understand the concepts I'm trying to teach, taking a test on paper with a number 2 pencil and not even color pictures is like going back in time. This test was originally written when you were about ten. I have kept it up-to-date conceptually, but for it to be truly up-to-date, a major overhaul of the test and the testing center are in order. The testing center administrator says that technical issues like security and authentication need to be worked out, and we just don't have a lab big, secure, or advanced enough to support the proper kind of testing. I don't see that happening very soon, either, with tight budgets and cramped space."

New Expectations

Due to technological advances, students' expectations have risen and educators' methods have changed, but assessment has utterly failed to keep pace. In this environment, students, educators, and especially researchers should be earnestly asking and investigating two primary technology-induced, assessment-related questions:

1. Should technologically literate students who live and learn in a multimedia environment at home, in the workplace, and at the university continue to have their learning assessed and measured through traditional pencil-and-paper-based methods?

2. Shouldn't our testing experts and researchers demand more alignment, or realignment, among learning objectives, instructional content, and assessment to ensure reliable and valid results now that technology has significantly altered at least one of the triad—presentation of the content?

The Gap

"Online," "technologically-integrated," "multimedia-based," and "e-learning" are familiar terms that describe a new approach to instructional delivery. Studies show that faculty and universities are quickly adopting these delivery modes. Analysts of this growth estimate that online learning is increasing 30–40 percent annually and that almost all higher education institutions will soon have online programs.

As an example of this growth, Brigham Young University (BYU) had only 50 courses online in the year 2000; as of fall term, 2003, it is expected that nearly one-third (2,000 of 6,500) of course sections will be at least partially online. In fall 2002, 81 percent of BYU students reported taking one or more of these courses, which commonly include quizzes and assignments administered only online. However, during this same period (2000–2003), there has been no corresponding change in the number of paper-based exams (800,000) administered by the university testing center. In fact, the center administered only 2,400 computer-based exams, 0.3 percent of the total, in the year 2002.

As impressive and overwhelming as the proliferation of e-learning at our universities and schools appears to the public and educators, it doesn't appear to impress students. Diana Oblinger's article in *Educause Review* highlights research revealing that students feel their

> teachers' use of technology is uninspiring; students consider themselves more Internet-savvy than their teachers; students report seeing better ways to use technology than do their teachers, … [and] their greatest use of technology is outside of school.[1]

Oblinger also noted a technological gap between high school students (Mil-

> **The academic discipline of assessment and measurement is built on two foundational pillars: reliability and validity**

lennial students) and eighteen- to twenty-two-year-old (Gen-X) college students wherein 94 percent of the so-called Millennial students "use the Internet for school research; … 70 percent use instant messaging to keep in touch; and … 56 percent prefer the Internet to the telephone."[2]

Millennial students might be disappointed by the absence—or the uninspiring use—of technology to enhance instruction, but they also must be disappointed by the lagging use of technology to assess and measure their learning. In an informal survey conducted recently (March 2003) at BYU, 70 percent of the Gen-X respondents—who already lag behind the Millennial students in comfort with and use of computer technology—either "strongly agreed" or "agreed" with the following: "As instructors use more computers in teaching, they should also test more using computers, rather than with paper and pencil."

While it is apparent that our current students, and especially our future students, expect more technology-enhanced learning and assessment, another, more significant, concern exists: digital-age students actually perform better when they take tests on computer. Boston College Researchers Russell and Haney reported that

> recent research shows that written tests taken on paper severely underestimate the performance of students accustomed to working on computers. The situation is analogous to testing the accounting skills of modern accountants, but restricting them to the use of an abacus for calculations.[3]

As disturbing as this finding might be, it should not be surprising that stu-

dents perform better on computer-aided than hand-written assignments or examinations. Students spend more time on computers than writing by hand (which also means that their psychomotor development is more suited to keyboarding and less to handwriting). For a student accustomed to keyboarding papers, projects, and assignments, the fatigue factor of using a number 2 pencil to hand-write essays, responses, and optical mark responses (OMR) for one to three hours would be explanation enough for poorer test scores. One testing center administrator responding to an informal survey posted on the National College Testing Association (NCTA) listserv made this comment: "We need to give the students the best opportunity possible to demonstrate their proficiency."

Alignment

The academic discipline of assessment and measurement is built on two foundational pillars: reliability and validity. The question we should ask about every test is, "How reliable and how valid is it?" Since no test will be completely reliable and perfectly valid, the test author's goal should always be to achieve as high a rate of predictability (reliability) and accuracy (validity) as possible.

Are paper-based testing practices reliable and valid for technologically sophisticated courses and students? One professor, quoting yet another professor, raised the reliability and validity question:

> As the environment changes for the delivery of instruction, it is important to reevaluate the ways in which we assess the learning outcomes for students using this new format and develop and apply assessment techniques that are more consistent with the learning environment.[4]

Consistency or predictability (reliability) between learning and testing environments that capitalize on students' technological adroitness will yield more accurate (valid) results. And while psychometricians identify and define different types of validity, this argument for more consistency or predictability—while not exclusive to the

other types of validity, especially the "construct validity" type—fits best under the category of "face validity": does the measurement fit, make sense, or look valid to an untrained observer? Wouldn't an observer from another world who visited our institutions of learning and witnessed the rich, multimedia learning environment and the technologically skilled students also expect the same kind of testing environment for these students? Can multimedia-rich instruction and paper-based assessment have any "face validity" when they are presented in different media?

In recent years there has been a call for "authentic assessment," a performance-type assessment that requires learners to demonstrate mastery of knowledge in either a real-world or simulated environment. Doesn't testing learners in the same way their future employers will expect them to apply what they have learned in a real-world, multimedia environment fit this definition of "authentic assessment?" Perpetuating the tradition of using two-dimensional, static, grayscale characters rather than adopting rich, multi-dimensional, animated, color, and interactive multimedia can only render our assessment inauthentic.

The real-life, just-in-time assessment contexts that technology makes possible are authentic. Maybe this is one reason students at BYU and many other educational institutions continue to rate exam performance very low. The question "Exams are good measures of my knowledge, understanding, or ability to perform" consistently receives the lowest rating of all 14 course characteristics (5.4 on a 7-point Likert Scale) on BYU's end-of-course evaluation instruments.

Causes and Solution

To solve a problem, there must really be a problem; and if there is really a problem, there must be causal agents. Assuming there really is a significant problem with our testing methods not aligning with both teaching methods and increasingly technologically literate students, what are some possible

causal agents? Is it resistance from the faculty or the testing-center administrators? Is it a lack of IT funding or strategic planning? Or is it a combination of these, plus other causes we haven't even identified yet?

Faculty. In 2000, BYU asked its faculty the following question as part of the institution-specific section of the National Survey of Student Engagement (NSSE) survey: "Would you consider having your classroom assessment administered online in the Testing Center?" Fifty-one percent (274) of the 539 responses said "yes." The follow-up question was this: "If yes, would you want to include more than a textual format, such as sound, graphics, videos, etc.?" to which 60 percent said "yes." These BYU responses would likely be even more significant if the proliferation of online courses (from 50 to 2,000) in the past three years is any indicator of growth and change. If it is, educators face a widening gap between objectives and content and assessment, and faculty seem willing to address the shortfall on their side.

Testing-Center Administrators. Eighty-four percent of NCTA-associated testing administrators said "yes" when asked, "Does your institution appear to have an increasing demand for proctored, computer-based, or online testing?" One respondent added, "The number of online classes is increasing. The majority of professors continue to use paper [and] pencil tests because we do not have a computer lab big enough to provide them a secure testing environment."

Is closing the gap between e-learning and e-assessment as simple as making multimedia-based exams available to faculty and students through secure computer labs with enterprise-wide testing systems? If so, why haven't educators already done it? Two testing-center administrators who responded to the NCTA-listserv survey seemed to cite logistics and lack of space as the primary obstacles: "I am turning down requests every week for online testing," and "We are refusing faculty the opportunity to test online because we can't

provide enough computers." As further evidence of this problem, the BYU testing center has 729 paper-based testing stations, but until recently had only 16 computer-based stations—which occupy more space per station than do the paper-based stations.

This gap between e-learning and e-assessment is clearly more complex than this article presents. However, testing-center administrators recognize that the problem is very real. The NCTA-listserv survey revealed that 67 percent of universities and colleges did not have (or respondents did not know they had) a strategic plan for e-assessment, and 89 percent called for additional research and discussion on this topic. Further, the implications of findings about the importance of this gap between teaching and testing contexts may be far-reaching. As one testing administrator commented, "This is a good topic for discussion. If we were to offer online, proctored testing for faculty, we would have to reorganize our entire operation."

Call for Change. Technologically literate students (and faculty) require an aligned teaching and testing environment to ensure reliable and valid measures of learning. The call for research, discussion, and strategic planning to close the gap between e-learning and e-assessment is not only timely, but urgent. *e*

Endnotes

1. D. Oblinger, "Boomers, Gen-Xers, and Millenials: Understanding the New Students," *Educause Review,* July/August 2003, pp. 36–47; <http://www.educause.edu/ir/library/pdf/erm0342.pdf>.

2. Ibid.

3. M. Russell and W. Haney, "Bridging the Gap Between Testing and Technology in Schools," *Education Policy Analysis Archives*, Vol. 8, No. 19, March 28, 2000; <http://epaa.asu.edu/epaa/v8n19.html>.

4. P. Frederick, "The Need for Alternative Authentic Assessments in Online Learning Environments," *Journal of Instruction Delivery Systems*, Vol. 16, No. 1, 2003, pp. 17–20.

Scott L. Howell (scott_howell@byu.edu) is Director of the Center for Instructional Design at Brigham Young University in Provo, Utah.