

Education Policy Analysis Archives

Volume 9 Number 5

February 14, 2001

ISSN 1068-2341

A peer-reviewed scholarly journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2001, the **EDUCATION POLICY ANALYSIS ARCHIVES**.

Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the [ERIC Clearinghouse on Assessment and Evaluation](#) and are permanently archived in *Resources in Education*.

How the Internet Will Help Large-Scale Assessment Reinvent Itself

Randy Elliot Bennett
Educational Testing Service
U.S.A.

Abstract

Large-scale assessment in the United States is undergoing enormous pressure to change. That pressure stems from many causes. Depending upon the type of test, the issues precipitating change include an outmoded cognitive-scientific basis for test design; a mismatch with curriculum; the differential performance of population groups; a lack of information to help individuals improve; and inefficiency. These issues provide a strong motivation to reconceptualize both the substance and the business of large-scale assessment. At the same time, advances in technology, measurement, and cognitive science are providing the means to make that reconceptualization a reality. The thesis of this paper is that the largest facilitating factor will be technological, in particular the Internet. In the same way that it is already helping to revolutionize commerce, education, and even social interaction, the Internet will help revolutionize the business and substance of large-scale assessment.

Whether for educational admissions, school and student accountability, or public policy, large-scale assessment in the United States is undergoing enormous pressure to change. This pressure is most evident with respect to high-stakes tests, like those used for grade promotion or college entrance. However, it is becoming apparent for lower-stakes survey instruments too, like the National Assessment of Educational Progress (NAEP) (e.g., Pellegrino, Jones, & Mitchell, 1999).

Several factors underlie the pressure to change. First, whereas our tests have incorporated many psychometric advances, they have remained separated from equally important advances in cognitive science, in essence measuring the same things in ever more technically sophisticated ways. Although decades of research have documented the importance of such cognitive constructs as knowledge organization, problem representation, mental models, and automaticity (Glaser, 1991), our tests typically do not account for them explicitly. As a result, our tests probably owe more to the behavioral psychology of the early 20th century than to the cognitive science of today (Shepard, 2000).

A second factor is the mismatch with the content and format of curriculum, a criticism more true of the developed ability tests commonly used in postsecondary admissions than of school achievement measures, but relevant to the latter too. The mismatch arises in part from the fact that the elemental, forced-choice problems dominating many tests are effective indicators of skills and abilities, and thus provide an efficient means for estimating student standing on those constructs. However, the mismatch becomes problematic because of the increasing attention being paid to test preparation. Although persistent direct training on these indicator tasks may increase test performance, it certainly is not the best way to improve construct standing. Further, it distracts attention from other, arguably more critical, learning activities (Frederiksen, 1984).

Differential performance of population groups is another factor. Because of the curricular mismatch, it is easy to blame group differences on purported bias in the test and more difficult to create a convincing defense than it would be if the tests were strongly linked to learning goals. In a high-stakes decision setting like admissions, tests become a lightning rod for the failure of schools and society to educate all groups effectively. With the potential elimination of affirmative action in university admissions, there is no politically acceptable choice but to reduce the role of such tests. California, Texas, Florida, and Pennsylvania are proposing to admit, or have begun admitting, all students with high-school rank above a certain point to their state higher education systems. At the same time, promotion tests tied to state curricular standards are being put into place to encourage schools to teach all students valued skills. Although in Texas one such test was challenged in court on the basis of differential performance, that challenge was rejected (Schmidt, 2000). This rejection suggests that when well-constructed tests closely reflect the curriculum, group differences should become more an issue of instructional inadequacy than test inaccuracy (Bennett, 1998).

As attention shifts to the adequacy of instruction, the ability to derive meaningful information from test performance becomes more critical. A weak connection between test and curriculum insures that the value of feedback for the examinee will be limited. Even for tests where the connection is stronger, feedback is still too often of marginal value, in part because of the additional cost and processing time that would be incurred. For achievement surveys like NAEP, which offer no information to individuals, schools, or districts, motivation to participate is undoubtedly diminished.

Finally, there is efficiency. Testing programs are expensive to operate. That

The Promise of New Technology

Radical improvements in assessment will derive from advances in three areas: technology, measurement, and cognitive science (Bennett, 1999). Of the three, new technology will be the most influential in the short term and, for that reason, I focus on it in this paper. New technology will have the greatest influence because it—not measurement and not cognitive science—is pervading our society. Billions of dollars are being invested annually to create and make commonplace powerful, general technologies for commerce, communications, entertainment, and education. Due to their generality, these technologies can also be used to improve assessment.

These technological advancements revolve primarily around the Internet. The Internet is (or will be) interactive, broadband, switched, networked, and standards-based. What does that mean?

- *Interactive* means that we can present a task to a student and quickly respond to that student's actions.
- *Switched* means that we can engage in different interactions with different students simultaneously. In combination, these two characteristics (interactive and switched) make for individualized assessments.
- *Broadband* means that those interactions can contain lots of information. For assessment tasks, that information could include audio, video, and animation. Those features might make tasks more authentic and more engaging, as well as allow us to assess skills that cannot be measured in paper and pencil (Bennett, Goodman, Hessinger, Liggett, Marshall, Kahn, & Zack, 1999). We might also use audio and video to capture answers, for example, giving examinees choice in their response modalities (typing, speaking, or, for a deaf student, American Sign Language).
- *Networked* indicates that everything is linked. This linkage means that testing agencies, schools, parents, government officials, item writers, test reviewers, human scorers, and students are tied together electronically. That electronic connection can allow for enormous efficiencies.
- Finally, *standards-based* means that the network runs according to a set of conventional rules that all participants follow. That fact permits both the easy interchange of data and access from a wide variety of computing platforms, as long as the software running on those platforms (e.g., Internet browsers), adheres to those rules too. (Note 1)

As an embodiment of these characteristics, what does the Internet afford? It affords the potential to deliver efficiently on a mass scale individualized, highly engaging content to almost any desktop; get data back immediately; process it; and make information available anywhere in the world, anytime day or night. Paper delivery cannot compete with this potential.

The Internet is, of course, not being built to service the needs of large-scale assessment. It is, instead, being built for e-commerce: to sell products and services over the web to consumers and to businesses directly. Coincidentally, the capabilities needed for e-commerce are essentially those needed for e-assessment:

- interactive (so that products can be offered and orders transacted),
- switched (so different business transactions can be conducted with different customers simultaneously),
- broadband (so that those offers can be as engaging and enticing as possible),

- networked (so that product offers, orders, shipping, inventory, and accounting can be integrated), and
- standards-based (so that everyone can get to it, regardless of computing platform).

Will we be able to count on continued investment in the Internet to support its use as a delivery medium? By any measure, the Internet and use of it, has grown dramatically, to say the least. As a communications medium, the Internet last year surpassed the telephone, with 3 billion email messages sent each day (Church, 1999). The number of unique URLs (web-page directory and subdirectory addresses) has grown from just under a billion in 1998 to a projected 3 billion in 2000 ("Big fish," 1999). In the United States, the percentage of homes with Internet access has increased from 26% in December 1998 to 42% in August 2000 (U.S. Department of Commerce, 2000). (Note 2) Worldwide, the number of users has grown from somewhere between 117 to 142 million in 1998 to about 400 million in 2000 ("Big fish," 1999; Global Reach, 2000; "How many online?", 2000). Finally, the number of host computers has gone from about 30 million to 75 million from January 1998 to January 2000 ("Internet domain survey host count," 2000). This phenomenal growth may slow as investment subsides from the speculative rates of the past few years. However, the vast size of the Internet and its user base constitute a critical mass that should continue to attract substantial capital.

For commerce, the promise of the Internet is all about being faster, cheaper, and better. Two "laws" of the digital era illustrate this promise. Moore's Law predicts the doubling of computational capability (specifically, at the level of the microchip) every 18 months. As Negroponte (1995) has explained, what filled a room yesterday is on your desk today and will be on your wrist tomorrow. Metcalfe's Law says that the value of a network increases by the square of the number of people on it. The true value of a network is, thus, less about information and more about community (Negroponte, 1995). One can see this effect clearly in eBay, the online auction broker (Cohen, 1999). Each new user potentially benefits every other existing user because every eBay member can be both buyer and seller. (Note 3) Metcalfe's law is playing out well beyond eBay. Online business-to-business auction brokers are appearing in a variety of industries, including natural gas, electricity, steel, and bandwidth (Friedman, 2000, pp. 386-387; Gibney, 2000).

Another illustration of this cheaper-faster-better result is the effect of the Internet on the traditional relationship between *richness* and *reach*, where *richness* is the depth of the interaction that a business can have with a customer and *reach* is the number of customers that a business can contact through a given channel. Traditionally, one limited the other. That is, a business could attain maximal reach but only limited richness. For example, through direct mail, broadcast, or newspaper ads a company could communicate with many people but have a meaningful interaction with none of them. Similarly, a business could attain maximal richness but limited reach. Via personal contact (e.g., door-to-door sales), very deep interactions can occur, but with only a relatively small number of people. What has the Internet done? It has transformed the relationship between richness and reach by allowing businesses to touch many people in a personalized but inexpensive way (Evans & Wurster, 2000). What does richness *with* reach make for? It makes for *mass customization*.

We can already see the effects in Dell Computer Corporation's business model. Customers can log onto Dell's Internet site (www.dell.com), choose from a menu of basic machine designs, and then configure a particular design to meet their needs. A second example is Radio.SonicNet (<http://radio.sonicnet.com/splash.asp>). Radio.SonicNet allows one to pick from a variety of music styles, choose artists within

that style, and indicate how frequently each artist should play. The end result is a radio station uniquely tuned to the individual and continually interesting; it always plays what you like but you never know exactly what it is going to play. As a final example, consider Customatix (www.customatix.com/customatix/common/homepage/HomepageGeneral.po), which allows you to design your own shoes using up to *three billion trillion* combinations of colors, graphics, logos and materials per shoe. You design them. They build them. And *nobody else* is likely to have exactly the same ones.

Reinventing Assessment

Reinventing the Business

There are two major dimensions to reinventing assessment. One is the business of assessment. This dimension centers on the core processes that define an enterprise. In many cases, those core processes can become many times more efficient because moving bits is faster and easier than moving atoms (Negroponte, 1995); that is, electronically processing information is far more cost effective than physically manipulating things.

For large-scale testing programs, some examples of the potential for electronic processing are in:

- developing tests, making the items easier to review, revise, and automatically morph into still more items (e.g., Singley & Bennett, in press) because the items themselves are digitally represented;
- delivering tests, eliminating the costs of printing, warehousing, and shipping tons of paper;
- presenting dynamic stimuli like audio, video, and animation, making the need for specialized testing equipment (e.g., audio cassette recorders, VCRs) obsolete (Bennett, Goodman, Hessinger, Ligget, Marshall, Kahn, & Zack, 1999);
- transmitting some types of complex constructed responses to human graders, removing the need to transport, house, and feed the graders (Odendahl, 1999; Whalen & Bejar, 1998);
- scoring other complex constructed responses automatically, reducing the need for human reading (Burstein et al., 1998; Clauser et al., 1997); and
- distributing test results, cutting the costs of printing and mailing reports.

To get a sense of how reinventing the business of assessment might affect testing organizations, take a look at reference book publishing, in particular the case of *Encyclopaedia Britannica* (Evans & Wurster, 2000; Landler, 1995; Melcher, 1997). *Encyclopaedia Britannica* was established in Scotland in 1768. It is the oldest and most famous encyclopedia in the English-speaking world. By 1990, its sales had reached \$650 million per annum. But then suddenly, *Britannica's* fortunes drastically changed. In 1996, the company was sold for less than half its net worth (i.e., the value of its assets, including its encyclopedia inventory, minus its liabilities). That same year, it eliminated its entire door-to-door North American sales force. By 1998, sales had fallen 80%. What happened?

What happened was that the reference book business was reinvented because of the emergence of new technology. At its peak, *Britannica* was a 32-volume set of books costing well over \$1,000. In 1993, Microsoft introduced *Encarta* on CD-ROM for under

\$100 and even though *Britannica* was much more comprehensive, the difference for most people wasn't worth an extra \$900+. Initially, *Britannica* did not respond as it didn't take the threat from *Encarta* seriously. But when it did respond, it did so ineffectively because *Britannica* wouldn't fit on a single CD-ROM and because the company's large sales force wasn't suited to selling software. But, ultimately, *Britannica* wasn't ready to cannibalize its existing paper business to enter this new electronic one.

Why is this story important? It's important because similar (though less extreme) scenarios are playing themselves out now in individual investing, book selling, travel planning, music distribution, long distance telephony, and even business-to-business transactions. (As to the last, Cisco Systems makes 90% of its revenue from business-to-business transactions done over the Internet [Cisco Systems, Inc., 2000]). These reinvention scenarios are forcing organizations—including some in educational assessment—to come quickly to grips with where new technology *will* and *will not* help core business processes.

As should be obvious, technology-driven changes in business processes can occur quickly and their consequences can be significant for the organizations that service a particular market. In fact, if radical and pervasive enough, process changes can force shifts in the substance of the business itself. So, although reinventing the business of assessment by incorporating technology into specific assessment processes is about trying to achieve the efficiencies needed to remain competitive today, reinventing the substance of assessment—most fundamentally, the reason we do it—is not about today. It's about tomorrow.

Reinventing the Substance

The populations seeking education are changing and so are their purposes for learning. At the college level, just 16% of students fit the traditional profile: 18-22 years old, full-time, on-campus resident (Levine, 2000a). This is not because fewer 18-22 year olds are going to college. It is because more adults are. The adult cohort is, in fact, the fastest growing segment in postsecondary education (Kerrey & Isakson, 2000). Working adults over age 24 constitute some 44% of college students ("Education prognosis 1999," 1999).

Why are so many adults returning to college? Over the past 25 years, employer demand in the U.S. has shifted toward higher educational qualifications, as indicated by an increasing premium paid for those with a college degree (Barton, 1999). But in addition to this rise in entry qualifications, the knowledge required to maintain a job in many occupations is changing so fast that 50% of all employees' skills are estimated to become outdated within 3-5 years (Moe & Blodget, 2000). Witness any job that requires interaction with information technology (IT), which is a growing proportion of jobs. In fact, by 2006 almost half of all workers will be employed by industries that are either major producers or intensive users of IT products and services (Henry et al., 1999).

So, more people want postsecondary education because they need to have it if they want to become—and stay—employed. And, more of these individuals are nontraditional students who may work, travel in their jobs, or have families. For these people, physically attending classes is not always feasible, let alone convenient. (Note 4)

This population's unmet educational need is increasingly becoming the target of distance learning. According to the National Center for Education Statistics, between fall 1995 and 1997-98, the percentage of higher education institutions offering distance learning courses increased by one-third (from 33% to 44%), and the number of course offerings and enrollments approximately doubled (Lewis et al., 1999). But although

many institutions have delivered distance learning via mail, radio, or television for years, this growth is not in those media. Rather, it is distance learning via the Internet that is booming. Among all higher-education institutions offering any distance learning, the percentage of institutions using asynchronous Internet-based technologies nearly tripled, from 22% in 1995 to 60% in 1997-1998. More recent data from Market Data Retrieval (MDR) confirm the trend ("Report: College Net use growing," 2000). MDR relates that, as of the 1999-2000 academic year, 34% of two- and four-year colleges offered accredited degree programs via computer, up from 15% the year before. As of 2000, U.S. institutions reportedly offered more than 6,000 accredited courses on the Web and, by 2002, over 2 million students will be enrolled, a tripling of the 1998 enrollment (Moe & Blodget, 2000).

At the same time, Internet-based distance learning is finding its way into high school. The need is generated by home-schooled students (of which there are over 1 million in the US), districts without a full complement of qualified teachers, and the children of migrant workers. So-called "virtual high schools" have emerged in Alabama, Arizona, California, Florida, Illinois, Indiana, Kentucky, Maryland, Massachusetts, Michigan, Missouri, Nebraska, New Mexico, and Utah (Carr, 1999; Carr & Young, 1999; Kerrey & Isakson, 2000). These programs can cross state lines, with offerings open to students regardless of residence. Of particular note is that both the University of Missouri at Columbia High School and the Indiana University High School have been granted accreditation by the North Central Association of Colleges and Schools (Carr, 1999). Accreditation means that students can apply course grades earned through these online institutions toward their high-school graduation. Both programs offer more than 100 high school courses.

The growth of Internet-based distance learning will have a significant impact upon traditional education. For one, it may threaten the existence of established institutions (Dunn, 2000; Levine, 2000b). Many in the private sector see education as a huge industry that produces mediocre results for a high cost. If the private sector can leverage new technologies, like distance learning, to deliver greater value, the institutions that dominate education today will not be the leaders tomorrow. The rapid growth of for-profit education companies (e.g., the University of Phoenix), and the seemingly endless creation of well-capitalized new ones (e.g., UNext, Caliber, KaplanCollege.com, University Access, K12), suggests that a serious challenge to the existing order is well underway. The gravity of the threat is evident in how non-profits have responded. Cornell University, Columbia University, the University of Maryland, and New York University, among others, have each announced their own for-profit distance learning subsidiaries (Carr, 2000a)!

A second reason that the growth of Internet-based distance learning will influence traditional education is that regardless of its impact on nonprofit institutions, the distance learning industry will produce sophisticated software that everyone can use, in school and out. Both Dunn (2000) and Tulloch (2000) suggest that this occurrence will blur the distinctions between distance learning and local education. APEX offers an example (<http://apex.netu.com/>). This company markets online Advanced Placement (AP) courses, targeting districts that want to offer AP but which do not have qualified teachers. Districts can, thus, use APEX offerings on site. (Note 5)

The considerable potential of online learning—local or distance—is reflected in a report to the President and Congress of the bipartisan Web-Based Education Commission (Kerrey & Isakson, 2000). The Commission reached the following conclusion:

The question is no longer *if* the Internet can be used to transform learning in new and powerful ways. The Commission has found that it can. Nor is the question *should* we invest the time, the energy, and the money necessary to fulfill its promise in defining and shaping new learning opportunity. The Commission believes that we should. (p. 134, italics in original)

If acted on, the consequences of this statement for assessment are profound. As online learning becomes more widespread, the substance and format of assessment will need to keep pace. Another quote from the Commission's report:

Perhaps the greatest barrier to innovative teaching is assessment that measures yesterday's learning goals... Too often today's tests measure yesterday's skills with yesterday's testing technologies—paper and pencil. (p. 59)

So, as students do more and more of their learning using technology tools, asking them to express that learning in a medium different from the one they typically work in will become increasingly untenable, especially where working with the medium is part of the skill being tested (or otherwise impacts it in important ways). Searching for information using the World Wide Web or writing on computer are examples. (Note 6)

These changes in learning methodology offer exciting possibilities for assessment innovation. On site or off, an obvious result of delivering courses via the Internet is the potential for embedding assessment, perhaps almost seamlessly, in instruction (Bennett, 1998). Since students respond to instructional exercises electronically, their responses can be recorded, leaving a continuous learning trace. Depending upon how the course and the assessment are designed, this information could conceivably support a sophisticated model of student proficiencies (Gitomer, Mislevy, & Steinberg, 1995). That model might be useful both for dynamically deciding what instruction to present next and for making more global judgments about what the student knows and can do at any given point.

In addition to assessment embedded in Internet-delivered courses, one can imagine Internet-delivered-assessment embedded in traditional classroom activity. Such assessment might take the form of periodically delivered exercises that both teach and test. In this scenario, the exercises would be standardized and performance might serve, depending upon the level of aggregation, to indicate individual, classroom, school, district, state, or national achievement. Thus, these exercises could serve summative as well as formative purposes and be useful to individuals as well as institutions. If the exercises were of high enough quality, such a model might improve the motivation to participate in voluntary surveys like NAEP.

There are, to be sure, many difficult issues:

1. How can we generate comparable inferences across students and institutions when variation in school equipment may cause items to display differently from one student to the next, potentially affecting performance?
2. How can we deliver assessment dependably given the unreliable nature of computers and the Internet, and the limited technical support available in most schools?
3. How might we make sense of the huge corpus of data that the electronic recording of student actions might provide?
4. How would student learning be affected by knowing that one's actions are being

recorded?

5. How can we prevent assessments that serve both instructional and accountability purposes from being corrupted by unscrupulous students or school staff?
6. How can we manage the costs of online assessment?
7. How can we assure that all parties can participate?

Let's, for the moment, turn to this last issue.

Are the Schools Ready?

A continuing concern with such reinvention visions is whether schools (and students) are ready technologically and, in particular, what to do about technology differences across social groups. The National Center for Education Statistics (NCES) reports that as of September 1999, 95% of schools were connected to the Internet, up from 35% in 1994 (NCES, 2000). Schools in *all* categories, (i.e., by grade level, poverty concentration, and metropolitan status), were equally likely to have Internet access. Further, most schools had dedicated lines: only 14% were using dial-up modem, a slower and less reliable access method.(Note 7)

Clearly many of these schools could have only a single connected machine and that machine could be the one sitting on the principal's desk. How many classrooms were actually wired? According to NCES (2000), as of September 1999, 63% of all instructional rooms had Internet access (up from 3% in 1994, a 20-fold increase in five years). The ratio of students to Internet-connected computers was 9:1, down from 12:1 only a year earlier. These are staggering numbers, for they imply that classrooms are connecting to the Internet at a very rapid rate.

This success is in no small part due to federal efforts. The government's *e-rate* program has been giving public schools and libraries discounts of up to 90% on phone service, Internet hook-ups, and wiring for several years ("FCC: E-rate subsidy funded," 2000). In total, the program has committed 3.65 billion dollars to over 50,000 institutions, helping connect more than one million public school classrooms (Kennard, 2000). In addition, 70% of the program's last round of funding went to schools in the lowest income areas.

However, even with these very significant efforts, there continue to be equity issues. As of September 1999, in high poverty schools, the ratio of students to Internet computers was 16 to 1. In low poverty schools, it was less than half that amount—7 to 1 (NCES, 2000).

What should we conclude? Certainly, with few exceptions, it would be impossible to deliver large-scale assessment via the Internet today. But the trend is clear: the infrastructure is quickly falling into place for Internet delivery of assessment to schools, perhaps first in survey programs like NAEP that require only a small participant sample from each school, but eventually for inclusive assessments delivered directly to the desktop. As evidence, witness the requests-for-proposals recently released by the state education departments of Oregon, Virginia, and Georgia for building Internet-delivered, state-assessment systems (Department of Education, 2000; Virginia Department of Education, undated, State of Georgia, 2001).

Assuming that every classroom is wired, will all students then have the technology skills needed to take tests on-line? Clearly, more students are becoming computer-familiar every day and developing such skills is a national educational technology goal (Riley, Holleman, & Roberts, 2000). But, as Negroponte (1995) suggests, computer familiarity is really the wrong issue. The secret to good interface

design is to make it *go away*. Thus, advances in technology will eventually eliminate the need to be computer familiar. After *nomadic* computing, which we are now entering with the proliferation of wireless Internet devices and personal digital assistants, comes *ubiquitous* computing (Olsen, 2000)—the embedding of new technology into everyday items. Inventions like "radio" paper (Gershenfeld, 1999, p. 18; Maney, 2000; "NCS secures rights," 2000) may allow students to interact with computers in the same way that they interact with paper today. Smart desks are another likelihood, in which case a test may be electronically delivered, quite literally, to every desktop.

In the U. S., then, we may see a future in which every classroom is wired and every student can easily take tests on line. What of the rest of the world? To be sure, the Internet is an American phenomenon. It derives from research sponsored by the Defense Department in the 1960's (Cerf, 1993). As a result of this history, the overwhelming majority of users were, until very recently, from our shores. At this writing, over 60% of Net users reside *outside* of the United States and the foreign growth rate now exceeds the domestic one ("How many online?", 2000; "U.S. dominance seen slipping," 2001).

The largest numbers of foreign Internet users are, of course, in developed nations. These nations have the telecommunications infrastructure and citizens with enough disposable income to afford the trappings of Internet use. But what about developing nations? Will they be left irretrievably behind? The challenges for these nations are undoubtedly great. Over time, however, we should see significant progress in building the infrastructure and the user base here too (Cairncross, 1997; Fernandez, 2000). This progress will occur for at least two reasons. First, the cost of technology has been dropping precipitously and, by Moore's law, will continue to decline. Further, because the future of computing is undoubtedly in wireless devices (Grice, 2000), a telecommunications infrastructure will be much cheaper to acquire than the land-lines of old. Second, as Metcalfe's law suggests, markets will become all the more valuable as they are interconnected. (Witness the global economy and the economic benefits resulting to nations from integration with it.) That developing nations join the e-commerce network means greater opportunity for all. It means more vendor choice for the people of developing nations; more opportunity for developed nations to serve these markets; and a new opportunity for third-world businesses themselves to compete globally. (Note 8)

The same holds true for assessment. The Internet will make it easier for developing nations to get access to assessment services from elsewhere and for those nations to distribute their own assessment services regionally or around the world. This ease of access and distribution should make it possible to form international consortia. Such consortia will be able to assemble technical resources that a single nation might not be able to acquire. In addition, those consortia may be able to purchase services from others more efficiently than nations could obtain individually. Finally, an electronic network should make it easier to participate in international studies, bringing the benefits of benchmarking to nations throughout the world.

But is Technology-Based Assessment Really Worth the Investment?

One of the largest instantiations of technology-based assessment to date is computer-based testing (CBT) in postsecondary admissions. As programs like the Graduate Record Examinations, the Graduate Management Admission Test, and the Test of English as a Foreign Language have found, CBT can be enormously costly. Being among the first large-scale programs to move to computer, they bore the brunt of creating the infrastructure for what was essentially a new business. The building of that

infrastructure was initiated in the early 1990's *before* test developers knew how to create tests for computer, *before* computers were widely available for individuals to take tests on, and *before* the Internet was ready to bring those tests to students. In essence, these programs needed to build both a factory to stamp out a new product and a new distribution mechanism. A first generation infrastructure now exists, but it is not yet optimized to produce and deliver tests as efficiently as possible. Right now, there's no question about it: for these programs, assessment by computer costs far more than assessment by paper.

If we have learned anything from the history of innovation, it is that new technologies are often initially far too expensive for mass use. That was true of the automobile, telephone service, commercial aviation, and the personal computer, among many other innovations. For example, in 1930 the cost of a three-minute telephone call from New York to London was \$250 (in 1990 dollars). By 1995, the cost had dropped to under \$1 (World Bank, 1995, cited in Cairncross, 1997, p. 28). As a second instance, when the IBM Personal Computer was introduced in 1981, it cost around \$5,000. At the time, the median family income in the United States was on the order of \$25,000, so that a computer cost about 20% of the average family's earnings—not very affordable. At this writing, the cost of a computer with many times greater capability is a little more than \$500 and the median income is closer to \$55,000. (Note 9) A computer now costs about 1% of average income. (Note 10)

When a promising new technology appears, individuals and institutions invest, allowing the technology to evolve and a supporting infrastructure to develop. Over the course of that development, failures inevitably occur. Eventually, the technology either dies or becomes commercially viable—that is, efficient enough.

So, who's investing in CBT? At this point, it's an impressive list including non-profit testing agencies, for profit-testing companies, school districts, state education departments, government agencies, and companies with no history in testing at all. The list includes ACT, the Bloomington (MN) Public Schools, CITO (the Netherlands), the College Board, CTB/McGraw-Hill, Edison Schools, ETS, Excelsior College (formerly Regents College), Harcourt Educational Measurement, Heriot-Watt University (Scotland), Houghton-Mifflin, Microsoft, the National Board of Medical Examiners, the National Institute for Testing and Evaluation (Israel), NCS Pearson, the Northwest Evaluation Association, the Oregon Department of Education, the Qualifications and Curriculum Authority (Great Britain), Thomson Corporation, the University of Cambridge Local Examinations Syndicate (UCLES), the U.S. Armed Forces, Vantage Technologies, and the Victoria (Australia) Board of Studies. These organizations are producing tests for postsecondary admissions, college course placement, course credit, school accountability, instructional assessment, and professional certification and licensure (see the Appendix for details.) In concert, they already administer something on the order of 10 million computerized tests each year. (Note 11)

Why are these organizations investing? I think it's because they believe that technology-based assessment will eventually achieve important economies over paper and that, fundamentally, assessment will benefit. But I also think it's because they don't want to become *Britannica*. That is, they see improvements in the business and substance of assessment which, if they fail to embrace, will lead them to the same fate as that encyclopedia publisher.

CBT as a Disruptive Technology

But as the case of admissions testing suggests, the road to improvement may be a

difficult one since CBT might not be a typical innovation. Christensen (1997) distinguishes between two types of innovation, called *sustaining* and *disruptive* technologies. Sustaining technologies enhance the performance of established products in ways that mainstream customers have traditionally valued. Historically, most technological advances in any given industry have been sustaining ones (e.g., in the personal computer industry, faster chips and bigger, higher-resolution monitors). Occasionally, disruptive technologies emerge. Companies introduce these technologies hoping their features will provide competitive edge. However, these features characteristically overshoot the market, giving customers more than they need or are willing to pay for. Thus, disruptive technologies result in *worse* product performance, at least in the near-term, on key dimensions in a company's established markets.

Interestingly, a few fringe customers typically find a disruptive technology's new features attractive. In these niche markets, such technology may thrive. If and when it advances to the level and nature of performance demanded in the mainstream market, the new technology can invade it, rapidly knocking out the traditional technology and its dependent practitioners. Remember *Britannica*.

CBT has many of the characteristics of a *disruptive* technology. Established testing organizations are applying it in their mainstream markets, most notably postsecondary admissions. This innovation was introduced, in good part, to provide competitive edge through features like the ability to take a test at one's convenience and to get score reports immediately. As it turned out, these features overshoot the market. At least initially, registrations for continuously-offered computer-based admissions tests mirrored those for fixed-date administrations, suggesting that scheduling convenience was not a highly valued feature in the market of the time. Moreover, examinees were dissatisfied with losing some of the features of paper exams, including the ability to proceed through the test nonlinearly, the option to review the scoring of items actually taken, and the low cost (Perry, 2000).

Although it encountered difficulty in the mainstream admissions testing market, CBT found more rapid acceptance in the niches. One example is information technology (IT) certification, which individuals pursue to document their competence in some computer-related proficiency. In 1999, over three million examinations in 25 languages were administered in this market (Adelman, 2000). Most of these tests were delivered on computer and most were offered on a continuous basis. Three delivery vendors provided the bulk of examinations: CAT, Inc. (a subsidiary of Houghton-Mifflin), Prometric (a subsidiary of Thomson Corporation), and Vue (a subsidiary of NCS Pearson). Together, these vendors operated some 5,000 testing centers in 140 countries. As of June, 2000, over 1.9 million credentials had been awarded, most for Microsoft or Novell technologies.

Why is the CBT of today so well suited to this market niche? Let's start by asking what features a testing product must have to succeed in this niche. First, it must be continuously offered because these test candidates build technology skill on their own schedules—at home or on the job, very often through books or online learning. These individuals want to test when they are ready, not when the testing companies are. Second, such a test must generally be offered on computer since technology use is the essence of the certification.

What are the financial considerations associated with serving this market? One consideration is whether the test fee can cover the cost of assessment. As it turns out, this market is less price-sensitive than postsecondary admissions. Why? With IT testing, employers pay the fee for over half the candidates (Adelman, 2000). In addition, certified employees command a substantial salary premium (4-14%), which makes

examinees more willing to absorb the higher fees that CBT currently requires. A second consideration is that security is not as critical as in admissions testing, so large item pools are not needed, reducing production cost. Lower security is tolerable because if an individual appears on the job with a dishonestly obtained credential but without the required skill, he or she will not last. Finally, test volume is self-replicating: there are many repeat test takers because information technology changes rapidly, so skills must be updated constantly. From an innovation perspective, then, IT certification may be one context in which the CBT of today can flourish and develop to better meet the needs of other assessment markets.

So why do industry leaders tend to fail with disruptive technology while fringe players succeed? Industry leaders often fail precisely because they attempt to introduce disruptive technologies into major markets before it's time (Christensen, 1997). Because niche markets are often too small to be of interest, leaders do not pursue those opportunities to refine the technology. Instead, they give up, having run out of resources or credibility. Making a disruptive technology work requires iteration and iteration means failure. Because they risk neither large resources nor reputations in the mainstream market, it is the fringe players who can fail early, often, and inexpensively enough to eventually challenge and overtake the industry leaders.

Toward the Technology Based Assessment of Tomorrow

Are there other niche markets in which CBT might evolve? One such niche may be online learning. If we believe the Web-Based Education Commission (Kerrey & Isakson, 2000), online learning will become a major enterprise, especially for the lifelong updating of skills. In this market, institutions will be less concerned with questions of who gets *in* and more with who gets *out*, and what it is they have to do to get out (Messick, 1999). Why? Because once hired, businesses are becoming more concerned with what employees know and can do, and less with where they went to school. Similarly, individuals are becoming more concerned with finding course offerings that meet their skill development goals and less with whether those offerings come from one institution or a half-dozen.

What's the assessment need? First, it is for *knowledge facilitation* and, second, for *knowledge certification*; that is, to help people develop their skills and then document that they've developed them. What's the assessment challenge? The challenge is to figure out how to design and deliver embedded assessment that provides instructional support and that globally summarizes learning accomplishment. In other words, the challenge is to combine richness with reach to achieve mass customization—use the Internet's ability to deliver the richness of customized assessment to reach a mass audience.

Can assessment be customized? In very rudimentary ways, it already is. Certainly, we can dynamically adapt along a global dimension, as is done in many of today's computerized tests. But as we move assessment closer to instruction, we should eventually be able to adapt to the interests of the learner and to the particular strengths and weaknesses evident at any particular juncture, as intelligent tutors now do (e.g., Schulze, Shelby, Treacy, & Wintersgill, 2000). Likewise, we should be able to customize feedback to describe the specific proficiencies the learner evidenced in an instructional sequence.

But perhaps the most far-reaching customization of assessment will come through modular online courses, whereby an instructor—or even a sophisticated learner—assembles a series of components into a unique offering. The Department of Defense (DOD) has taken a significant step through the Sharable Courseware Object

Reference Model (SCORM) (www.adlnet.org). SCORM is to embody specifications and guidelines providing the foundation for how DOD will use technology to build and operate the learning environment of the future. SCORM will allow mixing and matching of learning segments to create lower cost, reusable training resources. (Note 12) If embedded assessment can be built into course modules following a similar set of conventional specifications, the assessment too will be customized by default.

Conclusion

Whether for postsecondary admissions, school and student accountability, or national policy, large-scale assessment must be reinvented. Reinvention is not an option. If we do not reinvent it, much of today's paper-based testing will become an anachronism—"yesterday's testing technology," in the words of the Web-Based Education Commission (Kerrey & Isakson, 2000)—because it will be inconsistent with what and how students learn.

This reinvention must occur along both business and substantive lines. As educators, we often behave as if business considerations are unimportant, even distasteful. However, the business and substance of assessment are intertwined. Even for non-profit educational institutions—state education departments, federal agencies, schools, research organizations—providing quality assessment for a low cost matters. Using new technology to do assessment faster and cheaper can free up the resources to do assessment better.

We will be able to do assessment better because advances in technology, cognitive science, and measurement are laying the groundwork to make reinvention a reality. Whereas the contributions of cognitive and measurement science are in many ways more fundamental than those of new technology, it is new technology that is pervading our society. My thesis, therefore, is that new technology will be the primary facilitating factor precisely because of its widespread societal acceptance. (Note 13) In the same way that the Internet is already helping to revolutionize commerce, education, and even social interaction, this technological advance will help revolutionize the business and substance of large-scale assessment. It will do so by allowing richness with reach—that is, mass customization on a global scale—as never before. However, as the history of innovation suggests, this reinvention won't come immediately, without significant investment, or without setback. With few exceptions, we are not yet ready for large-scale assessment via the Internet (at least in our schools). However, as suggested above, this story is not so much about today. It really is about tomorrow.

Notes

This article is based on a paper presented at the annual conference of the International Association for Educational Assessment (IAEA), Jerusalem, May 2000.

I appreciate the helpful comments of Isaac Bejar, Henry Braun and Drew Gitomer on an earlier draft of this manuscript.

1. The Internet takes advantage of many such standards, including Internet Protocol (IP) for transmitting packets of information; Transmission Control Protocol (TCP/IP) for verifying the contents of those packets; HyperText Transfer Protocol (HTTP) for transferring web-pages; and HyperText Markup Language (HTML) and Extensible Markup Language (XML) for representing structured documents and data on the Web. XML provides a significant advance over HTML in that it

allows for the representation of unlimited classes of documents. Leadership in developing and implementing the many standards used by the Internet is provided by the World Wide Web Consortium (www.w3.org). For more on Internet standards, see their website or see Green (1996), who gives a more basic introduction.

2. According to Nielsen//NetRatings, 56% of U.S. households had Internet access as of November 2000 ("Internet access tops 56 percent," 2000).
3. And it works. eBay is reported to be the most successful company in cyberspace, with 22.5 million registered users and 2000 revenues of \$430 million (Cohen, 2001). Why? It has none of the costs of retailing: No buying, no warehousing, no shipping, no returns, no overstock.
4. A recent, but potentially significant, addition to this population is the U.S. Army. In July, 2000, Secretary of the Army, Louis Caldera, announced a 600 million dollar program to allow any interested soldier to take college courses over the Internet at little or no cost (Carr, 2000b).
5. A second, perhaps more interesting, example is Florida's Daniel Jenkins Academy, where students physically attend but take all academic courses *on-line* from *off-site* teachers (Thomas, 2000).
6. Russell has conducted several studies on the mismatch between learning and testing methods in writing (e.g., Russell & Plati, 2001). The repeated result is that the writing proficiencies of students who routinely use word processors are underestimated by paper-and-pencil tests.
7. The Teaching, Learning, and Computing—1998 survey provides similar data (Anderson & Ronnkvist, 1999). This survey, conducted using a national probability sample in Spring 1999, reports Internet access in 90% of schools and at least medium-speed, dedicated connections in 57%.
8. Developing a technology infrastructure and integrating into the e-commerce network may, in fact, help jump-start the growth required to deal with the serious problems of public health, education, and welfare that these countries typically face (Friedman, 2000).
9. The median income for a family of four in 1981 was \$26,274 (U.S. Census Bureau, 2001). For 1998, it was \$56,061.
10. Price and quality-adjusted data tell a similar story. In 1983, the quality-adjusted cost of a personal computer in constant 1996 dollars was \$1098 (D. Wasshausen, personal communication, April 13, 2000). By 1996, the cost of a PC, holding quality constant, was \$100, less than a tenth of the 1983 cost. By 1999, that quality-adjusted PC had further deflated to \$29.
11. I based this estimate on unduplicated volumes claimed by Thomson Prometric (www.prometric.com), Vantage Technologies (www.intellimetric.com/index.html), and the U.S. Armed Forces (A. Nicewander, personal communication, November 2, 2000). These three organizations *alone* claim some 8.5 million tests annually. These tests include both high-stakes and low-stakes assessments.
12. SCORM is being built upon the work of the IMS Global Learning Consortium (IMS) (www.imsproject.org/aboutims.html). IMS is developing open specifications for facilitating distributed learning activities such as locating and using educational content, tracking learner progress, reporting learner performance, and exchanging student records between administrative systems. Both IMS and SCORM incorporate XML (see note 1 above).
13. That the largest *facilitating* factor will be technological is not to say that we

should necessarily let technology drive the substance of assessment. We shouldn't.

References

- ACT and EDS alliance to expand the nation's testing and training opportunities. (1999, June 8). *ACT Newsroom* [On-line]. Available: www.act.org/news/releases/1999/06-08-99.html
- Adelman, C. (2000). A parallel postsecondary universe: The certification system in information technology. Washington, D.C.: Office of Educational Research and Improvement, U.S. Department of Education. Available: www.ed.gov/pubs/ParallelUniverse/
- Anderson, R. E., & Ronnkvist, A. (1999). The presence of computers in American Schools. Irvine, CA: Center for Research on Information Technology and Organizations. Available: www.crito.uci.edu/tlc/findings/computers_in_american_schools/
- Ball, S. (1999). Measurement and the culture of education: The story of VSAM. *Educational Measurement: Issues and Practice*, 18(2), 50-51.
- Barton, P. E. (1999). What jobs require: Literacy, education, and training, 1940-2006. Princeton, NJ: Policy Information Center, Educational Testing Service. Available: www.ets.org/research/pic
- Bennett, R. E. (1998). Reinventing assessment: Speculations on the future of large-scale educational testing. Princeton, NJ: Policy Information Center, Educational Testing Service. Available: www.ets.org/research/pic/bennett.html
- Bennett, R. E. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice*, 18(3), 5-12.
- Bennett, R. E., Goodman, M., Hessinger, J., Ligget, J., Marshall, G., Kahn, H., & Zack, J. (1999). Using multimedia in large-scale computer-based testing programs. *Computers in Human Behavior*, 15, 283-294.
- Big fish in a big pool. (1999). *TIME Digital*, December 2.
- Burstein, J., Braden-Harder, L., Chodorow, M., Hua, S., Kaplan, B., Kukich, K., Lu, C., Nolan, J., Rock, D., & Wolff, S. (1998). Computer analysis of essay content for automated score prediction (RR-98-15). Princeton, NJ: Educational Testing Service.
- Cairncross, F. (1997). *The death of distance: How the communications revolution will change our lives*. Boston, MA: Harvard Business School Press.
- Carr, S. (1999, December 10). 2 more universities start diploma-granting virtual high schools. *The Chronicle of Higher Education*, p. A49.
- Carr, S. (2000a, March 24). Cornell creates a for-profit subsidiary to market distance education programs. *The Chronicle of Higher Education*, p. A47.
- Carr, S. (2000b, August 18). Army bombshell rocks distance education. *The Chronicle of Higher Education*, p. A35.

- Carr, S., & Young, J. R. (1999, October 22). As distance learning boom spreads, colleges help set up virtual high schools. *The Chronicle of Higher Education*, p. A55.
- Cerf, V. (1993). How the Internet came to be. In B. Aboba (Ed.), *The online user's encyclopedia*. New York: Addison-Wesley. Available: <http://www.bell-labs.com/user/zhwang/vcerf.html>
- Christensen, C. M. (1997). *The innovator's dilemma: When new technologies cause great firms to fail*. Boston, MA: Harvard University Press.
- Church, G. J. (1999). The economy of the future? *TIME*, 154(14). Available: <http://www.time.com/time/magazine/article/0,9171,31522,00.html>
- Cisco Systems, Inc. (2000). *Discover all that's possible on the Internet: 2000 annual report*. San Jose, CA: Cisco Systems, Inc. Available: www.cisco.com/warp/public/749/ar2000
- Clauser, B. E., Margolis, M. J., Clyman, S. G., & Ross, L. P. (1997). Development of automated scoring algorithms for complex performance assessments: A comparison of two approaches. *Journal of Educational Measurement*, 34, 141-161.
- Cohen, A. (1999). The attic of e. *TIME*, 154(26). Available: <http://www.time.com/time/magazine/article/0,9171,36306-1,00.html>
- Cohen, A. (2001). eBay's bid to conquer all. *TIME*, 157(5), 48-51.
- Department of Education. (2000). Request for proposals for the technology enhanced student assessment system. Salem, OR: Department of Education. Available: www.ode.state.or.us/asmt/develop/rfptesa.htm
- Dunn, S. L. (2000). The virtualizing of education. *The Futurist*, 34(2), p.34-38.
- Early test prep. (1999). ABC News.com [On-line]. Available: <http://abcnews.go.com/sections/tech/DailyNews/testing991020.html>
- Education prognosis 1999. (1999, January 11). *Business Week*, 132-133.
- Evans, P., & Wurster, T. S. (2000). *Blown to bits: How the economics of information transforms strategy*. Boston, MA: Harvard Business School Press.
- FCC: E-rate subsidy funded at \$2.25 billion cap. (2000). *What Works in Teaching and Learning*, 32(8), p. 8.
- Fernandez, S. M. (2000). Latin America logs on. *TIME*, 155(19), B2-B4.
- Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39, 193-202.
- Friedman, T. L. (2000). *The Lexus and the olive tree: Understanding globalization*. New York: Anchor Books.
- Gershenfeld, N. (1999). *When things start to think*. New York: Holt.

- Gibney, Jr., F. (2000). Enron plays the pipes. *TIME*, 156(9), 38-39.
- Gitomer, D. H., Mislevy, R. J., & Steinberg, L. S. (1995). Diagnostic assessment of troubleshooting skill in an intelligent tutoring system. In P. D. Nichols, S. F. Chipman, R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 72-101). Hillsdale, NJ: Erlbaum.
- Glaser, R. (1991). Expertise and assessment. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 17-30). Englewood Cliffs, NJ: Prentice-Hall.
- Global Reach. (2000). Global Internet statistics (by language). Available: www.glreach.com/globstats/index.php3
- Green, C. (1996). An introduction to Internet protocols for newbies. Available: www.halcyon.com/cliffg/uwteach/shared_info/internet_protocols.html
- Grice, C. (2000). Wireless handhelds will rule the day, PC execs predict. *CNET News.com* [On-line]. Available: <http://news.cnet.com/news/0-1004-200-1560446.html>
- Henry, D., Buckley, P., Gill, G., Cooke, S., Dumagan, J., Pastore, D., & LaPorte, S. (1999). The emerging digital economy II. Washington, D.C.: U.S. Department of Commerce. Available: www.ecommerce.gov/ede/ede2.pdf
- How many online? (2000, December 20). Nua Internet Surveys. Available: www.nua.ie/surveys/how_many_online/index.html
- Internet access tops 56 percent in U. S., according to Nielsen//NetRatings. (2000, December 18). Available: http://209.249.142.22/press_releases/PDF/pr_001215.pdf
- Internet domain survey host count. (2000). Internet Software Consortium. Available: www.isc.org/ds/hosts.html
- Kennard, W. E. (2000, January). E-rate: A success story. Presentation at the Educational Technology Leadership Conference—2000, Washington, D.C.
- Kerrey, B., & Isakson, J. (2000). The power of the Internet for learning: Moving from promise to practice. (Report of the Web-based Education Commission). Washington, D.C.: Web-Based Education Commission. Available: <http://interact.hpcnet.org/webcommission/index.htm>
- Landler, M. (1995, May 16). Slow-to-adapt *Encyclopaedia Britannica* is for sale. *New York Times*, D1, D22.
- Levine, A. (2000a, March). The remaking of the American university. Paper presented at the Blackboard Summit, Washington, D. C.
- Levine, A. (2000b, March 13). The soul of a new university. *New York Times*, p. 21.
- Lewis, L., Snow, K., Farris, E., Levin, D., & Greene, B. (1999). Distance education at postsecondary education institutions: 1997-1998 (NCES Statistical Analysis Report 2000-013). Washington, D.C.: National Center for Education Statistics. Available:

<http://nces.ed.gov/pubs2000/2000013.pdf>

Maney, K. (2000). E-novel approach promises new chapter for book lovers. *USA Today*, 18(169), 8A-9A.

Melcher, R. A. (1997). Dusting off the *Britannica*: A new order has digital dreams for the august encyclopedia. *Business Week Online*. Available: www.businessweek.com/1997/42/b3549124.htm

Mendels, P. (1999). The leading issues of '99? Wired schools and accreditation. *The New York Times On the Web* [On-line]. Available: www.nytimes.com/library/tech/99/12/cyber/education/29education.html

Messick, S. (1999). Technology and the future of higher education assessment. In S. Messick (Ed.), *Assessment in higher education: Issues of access, student development, and public policy* (pp. 245-254). Hillsdale, NJ: Erlbaum.

Moe, M. T., & Blodget, H. (2000). *The knowledge web: People power—Fuel for the new economy*. San Francisco: Merrill Lynch.

National Center for Education Statistics. (2000). Stats in brief: Internet access in US public schools and classrooms: 1994-99. Washington, DC: US Department of Education, Office of Research and Improvement.

NCS secures rights to iPaper electronic technology in testing and education market. (2000, July 11). Minneapolis, MN: National Computer Systems (NCS). Available: www.ncs.com/ncscorp/top/news/000711.htm

Negroponete, N. (1995). *Being digital*. New York: Vintage.

Odendahl, N. (1999, April). Online delivery and scoring of constructed-response assessments. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Olsen, F. (2000, February 18). A UCLA professor and net pioneer paves the way for the next big thing. *The Chronicle of Higher Education*, 46.

Pellegrino, J. W., Jones, L. R., & Mitchell, K. J. (1999). *Grading the nation's report card*. Washington, D.C.: National Academy Press.

Perry, J. (2000). Digital tests spark controversy: Critics say revamped exams limit the options to challenge a score. *Online US News* [On-line]. Available: www.usnews.com/usnews/edu/beyond/grad/gbgre.htm

Poised to go global: Accuplacer online sales soar. (2000, April). *The Bulletin Board*, 5(9), 5.

Report: College Net use growing. (2000, March 16). *USA Today.com* [On-line]. Available: www.usatoday.com/life/cyber/tech/cth566.htm

Riley, R. W., Holleman, F. S., & Roberts, L. G. (2000). e-Learning: Putting a world-class education at the fingertips of all children (The national educational

- technology plan). Washington, D.C.: U.S. Department of Education. Available: www.ed.gov/Technology/elearning/e-learning.pdf
- Russell, M., & Plati, T. (2001). Effects of computer versus paper administration of a state-mandated writing assessment. *Teachers College Record*. Available: www.tcrecord.org/Content.asp?ContentID=10709 .
- Schmidt, P. (2000, January 21). Judge sees no bias in Texas test for high-school graduation. *Chronicle of Higher Education*, p. A27.
- Schulze, K. G., Shelby, R. N., Treacy, D. J., & Wintersgill, M. C. (2000, April). Andes: A coached learning environment for classical Newtonian physics. In *Proceedings of the 11th International Conference on College Teaching and Learning*, Jacksonville, FL. Available: www.pitt.edu/~vanlehn/icctl.pdf
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Singley, M. K., & Bennett, R. E. (in press). Item generation and beyond: Applications of schema theory to mathematics assessment. In S. Irvine & P. Kyllonen (Eds.), *Item generation for test development*. Hillsdale, NJ: Erlbaum.
- State of Georgia. (2001). Request for proposal number 41400-026-0000000031. Available: <http://www2.state.ga.us/Departments/doas/procure/rfp/rfp-41400-026-0000000031.doc>
- Thomas, K. (2000, April 6). One school's quantum leap. *USA Today*. 1A. Available: www.usatoday.com/usatoday/20000406/2117463s.htm
- Tulloch, J. B. (2000). Sophisticated technology offers higher education options. *T.H.E. Journal* [On-line]. Available: www.thejournal.com/magazine/vault/A3165.cfm .
- U.S. Census Bureau. (2001). Median income for 4-person families, by state. Available: www.census.gov/ftp/pub/hhes/income/4person.html
- U.S. Department of Commerce. (2000). Falling through the Net: Toward digital inclusion. Available: www.esa.doc.gov/ftn00.pdf
- U.S. dominance seen slipping in Internet use, commerce. (2001). Cyberatlas: The Big Picture Geographics. Available: http://cyberatlas.internet.com/big_picture/geographics/article/0,,5911_377801,00.html
- Virginia Department of Education. (Undated). Demonstrating success: A statewide web-based Standards of Learning technology and on-line testing initiative (Request for proposal # RFP-WEB2000). Richmond, VA: Virginia Department of Education. Available: www.pen.k12.va.us/VDOE/Technology/soltech/rfp/rfpweb2000.pdf
- Whalen, S. J., & Bejar, I. I. (1998). Relational databases in assessment: An application to online scoring. *Journal of Educational Computing Research*, 18, 1-13.

Appendix: Some Organizations Investing in Computer-Based Testing

ACT, Inc. In partnership with EDS, ACT, Inc. is establishing a nationwide network of electronic testing and training centers. These centers will provide computer-delivered certification and licensure tests for the trades and professions; a computerized measure of workplace skills to guide training decisions; and computerized educational and career guidance. More than 250 ACT Centers are expected to be operational by the end of 2001 ("ACT and EDS," 1999). ACT also offers a computerized placement test for post-secondary institutions to use in determining whether entering students need assignment to remedial or developmental courses in mathematics, reading, writing, and English-as-a-second-language (www.act.org/compass/).

Bloomington (MN) Public Schools. This district was reportedly the first in the US to do its math and reading testing exclusively via computer ("Early test prep," 1999). Bloomington uses an intranet-delivered computer-adaptive test designed by the Northwest Evaluation Association (see entry below) (www.bloomington.k12.mn.us/Staff_Resources/Office_of_Research_and_Evaluat/CALT_Technical_Description/calt_technical_description.htm).

CITO. CITO, the measurement organization of the Netherlands, has developed a computerized adaptive test, WisCat, for placement in adult education. WisCat is used by approximately half the vocational training institutes in the Netherlands (Verschoor, personal communication, November 7, 2000).

College Board. The College Board offers Accuplacer, an adaptive placement test that can be delivered over the Internet for use in postsecondary institutions (www.collegeboard.org/accuplacer/html/accupla1.html). Last year, over 2 million exams were administered ("Poised to go global," 2000), probably making Accuplacer the largest volume CBT in the world. By July 2001, the Board will also be offering its entire College Level Examination Program (CLEP) on computer: over 30 tests designed to allow individuals to get college credit for knowledge gained outside of school (www.collegeboard.com/clep/clepctr/html/tc001.html).

CTB/McGraw-Hill. This company offers a PC version of the Test of Adult Basic Education, a measure of reading, mathematics, language, and spelling skills used in adult literacy programs (www.ctb.com/products_services/tabe/index.html).

Edison Schools. This for-profit company manages 113 public schools with a total enrollment of 57,000 students. Edison recently introduced its Benchmark Assessment System, designed to provide teachers with ongoing, instructionally relevant information about the progress of their 2nd to 8th grade students. These computerized assessments in reading, math, writing, and language arts will be administered over 1 million times during the 2000-2001 academic year (www.intellimetric.com/when.newstoday0.html).

Educational Testing Service (ETS). In the 1999-2000 year, ETS administered over a million tests on computer for the GRE, GMAT, and TOEFL programs. In addition, a variety of licensure and certification examinations were given through ETS' Chauncey Group International subsidiary (www.ets.org/cbt/index.html). A second subsidiary, ETS Technologies, markets automated scoring services for computer-delivered writing tests

The World Wide Web address for the *Education Policy Analysis Archives* is epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, [Gene V Glass](mailto:glass@asu.edu), glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

[Michael W. Apple](#)

University of Wisconsin

[John Covalesskie](#)

Northern Michigan University

[Sherman Dorn](#)

University of South Florida

[Richard Garlikov](#)

hmwkhelp@scott.net

[Alison I. Griffith](#)

York University

[Ernest R. House](#)

University of Colorado

[Craig B. Howley](#)

Appalachia Educational Laboratory

[Daniel Kallós](#)

Umeå University

[Thomas Mauhs-Pugh](#)

Green Mountain College

[William McInerney](#)

Purdue University

[Les McLean](#)

University of Toronto

[Anne L. Pemberton](#)

apembert@pen.k12.va.us

[Richard C. Richardson](#)

New York University

[Dennis Sayers](#)

Ann Leavenworth Center
for Accelerated Learning

[Michael Scriven](#)

scriven@aol.com

[Greg Camilli](#)

Rutgers University

[Alan Davis](#)

University of Colorado, Denver

[Mark E. Fetler](#)

California Commission on Teacher Credentialing

[Thomas F. Green](#)

Syracuse University

[Arlen Gullickson](#)

Western Michigan University

[Aimee Howley](#)

Ohio University

[William Hunter](#)

University of Calgary

[Benjamin Levin](#)

University of Manitoba

[Dewayne Matthews](#)

Western Interstate Commission for Higher
Education

[Mary McKeown-Moak](#)

MGT of America (Austin, TX)

[Susan Bobbitt Nolen](#)

University of Washington

[Hugh G. Petrie](#)

SUNY Buffalo

[Anthony G. Rud Jr.](#)

Purdue University

[Jay D. Scribner](#)

University of Texas at Austin

[Robert E. Stake](#)

University of Illinois—UC

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia
Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación
Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de
México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de
México
humberto@servidor.unam.mx

Daniel Schugurensky
(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de
México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do
Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia
e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu