Technology for Assessment:
Tackling the Policy Issues

Conference Summary

Mary Axelson

Edited by
Sheila McEntee

In June 2004, in conjunction with the Council of Chief State School Officers' (CCSSO) 34th national conference on large-scale assessment, the Institute for the Advancement of Emerging Technologies in Education (IAETE) at AEL hosted a daylong symposium to discuss critical policy issues related to technology-supported assessments. Held in Boston, Massachusetts, the symposium was the last of three such forums sponsored by IAETE since 2002 to explore technology's potential contributions to assessment systems. Where earlier symposia gathered education practitioners and researchers, this event brought the issue to policy influencers and implementers. The unique perspectives and priorities of these three groups were represented in each discussion via the presenters and attendees. The format encouraged audience participation through focused discussion sessions and interaction with the panelists.

AEL holds a national leadership designation in the area of new and emerging technologies for the Regional Educational Laboratory Network, which is sponsored by the U.S. Department of Education's Institute of Education Sciences (IES). IAETE carries out this work for AEL, and it is in this visionary capacity that symposia participants gathered.

The kinds of technologies that can support assessment have grown substantially. There are technologies for measuring student knowledge, managing and interpreting data, protecting data on student performance, and distributing and displaying results. There are online assessments that look very much like familiar paper-and-pencil assessments, but with the benefit of faster results. All represent experiments in entirely new assessment offerings.

Cohosted by CCSSO, IAETE's third symposium focused on the potential to develop assessments that measure depth and maturity of knowledge, rather than discrete bits of information. Behind this concept are two significant publications from the National Research Council: *How People Learn* and *Knowing What Students Know*. Central to both are research findings in cognitive science that suggest that what a student knows is less significant than how he or she can make use of knowledge. Experts, for example, organize knowledge differently than do novices. Technologies that could demonstrate how a student is organizing information could create an assessment revolution.

Discussions at the symposium expanded the scope of what could be considered assessment technologies. For example, participants suggested that while putting existing multiple-choice tests online may not be as glamorous as, say, assessment via concept mapping or simulation software, the task of replicating familiar assessment forms in online environments is vital to the

evolution of assessment technologies. So, too, are the early experiences of creating new assessment types that use technology to assess skills and knowledge in ways that cannot be accomplished by more traditional means.

Two clear purposes of IAETE's three-part symposia series were to address the requirements of No Child Left Behind (NCLB) and priorities for the use of technology by schools as outlined by then-Secretary Rod Paige of the U.S. Department of Education. In the first two symposia, participants cited the limitations of the large-scale assessment systems currently being used to meet NCLB accountability requirements, saying they often provided too narrow a picture for too great a purpose. In contrast, attendees of the final forum, who inform policy development and implement policies, frequently voiced support of the scope of assessment data required by NCLB. Their interests focused not on radical change but on realistic, immediate improvement to assessments.

A striking unanimity of purpose ruled the day, perhaps best expressed by panelist Bob Olsen, who said, "It's all about the student. It's all about helping the teacher help the student." Though school measurement by large-scale assessment is a priority for this group, they viewed their work as assembling a mix of assessments for different purposes. Repeatedly, speakers and participants said that classroom assessments carry the greatest value, and that immediate feedback on student performance is their greatest hope from technology. The importance of educating classroom teachers about assessment came to the fore, as did the promise and surprising difficulties of technology-based testing for special needs students.

The day's conversations were structured around two perspectives common to much of the work at IAETE: "Vision," which anticipates what can be, and "Leadership," which focuses on the here and now of best practices. A group of panelists addressed each perspective.

Vision

- **Joe Kitchens,** Superintendent, Western Heights Public Schools, Oklahoma City, Oklahoma
- **David J. Harmon,** Administrator, Walton High School, Cobb County Schools, Georgia
- **Kevin Ruess,** Founder and Principal, Ludactica, an instructional design firm with a focus on playful learning
- **Bob Olsen,** Director of Research and Assessment, Bend-La Pine School District, Oregon

Managing Data in Oklahoma's Western Heights School District

For teachers to be empowered decision makers, they have to have access to data, and they have to have access to data in real time. We can't have systems where we are waiting three months for data to arrive, because a quarter of the school instructional year is then gone. Learning is a dynamic thing that is happening every day, every hour, every minute.

—Joe Kitchens

"Cat herding." That is how Joe Kitchens describes what it is like to manage school data. It is a telling description, coming from someone whose efforts are among the most sophisticated in the country. Kitchens, superintendent of the 3,100-student, seven-school district of Western Heights, Oklahoma, has relentlessly pursued the goal of improving student learning with data on student performance. He does this with technology. He has strived to put real-time data in teachers' hands immediately after an assessment—at least within a day—because, he said, "this is a business where every minute counts." His small district has also invested in a statistician to help wrangle the numbers into something of value in the classroom.

In Kitchens' view, managing student performance data is a district responsibility—largely because it is unwieldy to send all the data back and forth between the district and the state. His district, for example, has "300 teachers and administrators who constantly need to review the data from 5,000 course sections in a year," he said. That creates 60,000 points of contact each year." That *is* cat herding.

Kitchens began the panel discussion by clearly stating his support of NCLB:

I want to say this. I am a proponent of NCLB. I believe very strongly that the accountability and assessment requirements of NCLB are a good thing. But I also

believe that accountability for the sake of accountability is not a good thing. We have to actually use those assessments in a way that promotes student learning.

Kitchens believes we need to "redesign and retool our system." Most important to a redesign of the information management system, in his experience, are some standard definitions and presentations of assessment data. Indeed, he is on a quest for a common vocabulary. XML, a computer language that enables the efficient identification and labeling of data types, would be preferred for this purpose, Kitchen says, but no testing company has ever offered him results in XML. He also wants to have data ready to run with demographic information required by NCLB. Ideally, all these descriptions would be synchronized with content descriptors for the instructional management system. "This should be very definitive," says Kitchens, "and technology will allow us to get very definitive." The U.S. Department of Education and assessment providers, he says, might best spearhead such an effort.

Georgia Schools Increase Use of Web-Based Item Bank

We were smart enough to know that folks in schools and districts needed to have access to Web-enabled assessment well before it "counted" in an accountability sense.

—David J. Harmon

The goal of leveraging emerging assessment and accountability requirements to improve teaching and learning encouraged the state of Georgia to create a Web-based item bank for its criterion-referenced competency tests (CRCT) (http://www.doe.k12.ga.us/curriculum/testing/crct.asp). Specifically, the state expected to measure higher cognitive skills and processes not currently assessed, reduce turnaround time on score reports, reduce labor associated with paper-and-pencil assessments, and ultimately save money. Now, the state's 30,000-item bank is organized into three secure levels of access that serve distinct assessment purposes. Level 1 is available to students and parents for self-assessment, remediation, or enrichment. Level 2 is accessible to teachers for creating classroom tests upon completion of instructional units or sequences of instruction. Level 3 items are reserved for end-of-year high-stakes tests.

In reviewing how Georgia got to this point, David J. Harmon, the Cobb County Schools administrator who led the project through the legislative process and through implementation in schools, said Georgia started out with an RFP that bundled the technology and assessment portions of the item bank together. Officials refined their picture of what they wanted as they listened to proposal presentations.

The initial scope was to provide the instructional items (levels 1 and 2) for every grade, and they pushed the contractor hard to achieve this by the 2002-2003 school year. The beginning, said Harmon, was a "disaster." Student data, which had to be entered by hand, would disappear. But the system started building some credibility by the end of the year, culminating with a record of 31,000 "test events" (a test developed or administered) recorded in a single day. In all, there were approximately 1.3 million test events that year. That number climbed more than 50 percent to 2 million test events in the 2003-2004 school year. Harmon believes the instructional component of the bank (level 2), adds great value to the classroom and builds support for the eventual full utilization of the Web-based item-banking system.

The multiple-choice items at levels 1 and 2 are scored electronically. However, constructed-response items, performance assessments, and problem simulations require human scoring, using scoring guidelines and rubrics. Students judge their own work when using level 1; teachers score these items when administered from level 2.

Georgia has not yet implemented level 3 for high-stakes accountability, and there are no plans for high-stakes testing in the 2004-2005 school year. They have, however, put some new state tests online. Georgia reform legislation calls for end-of-course tests in eight high school subjects. These were first administered in the 2003-2004 school year, and 30 schools implemented the online version of at least one of the end-of-course, contractor-scored tests. Georgia will continue to train and build the infrastructure for the assessment process.

Security issues, Harmon said, are holding back high-stakes tests online. For example, if there is only one version of the test and kids are in a lab, they can look around at other screens. He also remains dissatisfied with accommodations for special education students, though he believes they will be a benefit of online testing down the road.

Harmon discussed his hopes for the future of technology-based assessment:

> Putting a multiple-choice test online is not that exciting. It's difficult, but not terribly exciting. So, what we wanted to do was to tap into higher levels of cognition, some problem solving. We wanted to do things that you can't necessarily do or measure quite as well with paper and pencil.

Defining purpose, said Harmon, also influenced how items were constructed. Teaching and learning, not just accountability, were their purpose. In Georgia, they ran focus groups through-

out the state, during which they asked teachers to describe components of a testing program that would be most useful in classrooms. As Harmon said, "Whether it's driven by NCLB or Georgia reform, it's still about the classroom and the child's learning." When talking to the state department, however, Harmon found himself highlighting benefits such as cost savings from decreased turnaround time and the reduction of labor requirements for the logistics of testing.

Instructional Design and Gaming Meet Assessment

Technology is not about shifting media; if you have taken your multiple-choice test and moved it from bubble sheets to a Web page, you have not done anything new. You have made it easier to collect data, without a doubt, and that is a huge gain. But that is not a gain in assessment per se; it is a gain in the management of assessment data.

—Kevin Ruess

Panelist Kevin Ruess brought an instructional design perspective to the conversation. As founder and principal of Ludactica, LLC, he leads an instructional design company that focuses primarily on the use of games in K-12 settings. Said Ruess, "I have thought mostly about games for instruction, for learning. I hadn't really thought about them for assessment until I was asked to sit on this panel. There are some really interesting possibilities here."

Ruess works with multiplayer games. He and Christopher Dede were principal investigators of the Multi-User Virtual Environment Experiential Simulators (MUVEES) project, a research effort funded by the National Science Foundation and developed in partnership with Harvard University. Ruess and Dede created a multiplayer mystery game in which students figure out why the residents of a late nineteenth century American town are ill. To advance in the game, players must both learn and share what they know with others.[1]

Ruess identified three key ways technology contributes to assessment:

1. enhancing learning
2. shortening the feedback loop
3. guiding remediation

---

[1] Dr. Dede explained this work at the initial symposium in this series

"Interestingly," observed Ruess, "games do a lot of these things very well." Games are, for example, filled with decision points and other moments that are, essentially, assessments. The data that are collected for tracking the game could also provide assessment insight. Furthermore, games provide transparent assessment with immediate feedback. His vision is to design an assessment that doubles as a learning experience. "If a learner doesn't already know it when they are taking the test," he asks, "will your test help them learn it *while* they are taking the test?"

Ruess discussed two principles that drive instructional designers. Adoptability was first. "Can it actually be used in the environment for which it is intended?" he asked. To design effective, adoptable assessments, designers must understand the reality of the school day. For example, it is inevitable, he said, that an assembly will be called on the day a 45-minute assessment is planned, thus limiting class time to 32 minutes. So, he is now working on 10-minute games because they fit better into hectic school periods.

Second on his instructional design wish list was an endogenous, as opposed to an exogenous, environment. Ruess explained it this way:

> So let's take an example of a [popular math arcade-style] type game. . . . There is no relationship between shooting down asteroids and being able to add or subtract. That is an exogenous structure. An endogenous structure would be a situation where the structure itself is exactly what you are trying to do in the games. . . . Most assessments that I can think of are exogenous structures. They are not related to what you are trying to do; they are related to measuring the learning as opposed to the learning itself, and that, I think, is the real challenge before us.

Oregon's Online Testing Experience and Keeping a Kid's View in Mind

My job as an assessment professional is to provide to people who can use that information, information about what a kid knows and can do, and to do it as efficiently and accurately as possible.

—Bob Olsen

Bob Olsen is director of research and assessment for the Bend-La Pine School District in central Oregon. For four years, he served as director of the Technology Enhanced Student Assessment (TESA) Systems, an Oregon Department of Education project that delivers the state's testing program to students via the Web. The system gives kids results in a mouse click, and that,

says Olsen, "is absolutely the most important thing you can do. It's much more important that the kid knows it than that the teacher knows it."

As Olsen considered policy issues for this forum, he first turned to standards. "Isn't that where assessments begin?" he asked. He believes we need practical people to define standards, not people who are passionate about the subject area. Offering writing as an illustration, he explained, "Today's writing to survive, to prosper, looks nothing like the writing we were teaching 20 years ago. Let people in the real world tell us how good our kids need to be."

Olsen also spoke of the importance of professional development, a recurring issue at the forum. "Let's help and let's train classroom teachers to be assessment professionals," he said. "The last time I looked, admittedly a decade ago, less than 11 percent of the teacher preparation institutions in the country required that even a topic of assessment be presented to teacher education candidates. Not a course, just a topic in a course."

The "biggest point" Olsen wished to bring to the conversation was the need for student advocates. Making a distinction between the customer and the consumer, Olsen asked, "Who speaks for the consumer? Who speaks for the kids?" Praising Dick Baldwin, one of Olsen's former supervisors, for saying, "Every once in a while you have to go get kid on you or you can't do your job," Olsen recommended a school system similar to those in Britain and Australia. There, he explained, "they rotate practitioners into their departments of education and rotate people in the department of education out into the field." Olsen added that he has lobbied for the system with "zero success." TESA, he said, "kept kid on it" by treating its work much like a commercial game development company would. "State boards of education," advised Olsen, "could hire independent researchers to speak for the kids. It would give them a voice that is not presently heard."

People "with kid on them," contended Olsen, know that children do not discuss test items on the playground and that a test item will not follow a kid from state to state. "One of the things that would be most exciting," he concluded, "would be for states to begin sharing. States have at their disposal tremendous resources, in terms of item banks, in terms of the technology to apply those item banks." All, he said, are overzealously protected.

The Audience Responds

Proof of the overly protective attitude identified by Olsen was evident when the audience was invited to ask questions. One attendee explained that principals were not giving teachers

passwords to student information sites as expected. Kitchens recommended that schools and districts structure passwords with a Web services model that uses an active directory. The system currently used by Kitchens' staff provides secure levels of access to Web-based services. It has helped his staff to overcome a reluctance to incorporate these services into teaching, learning, and school management.

Jan Barth of the West Virginia Department of Education, a panelist later in the day, asked Kitchens, "How did you get to the point of people understanding that the summative assessment is something you use, but the high-yield ticket item is the classroom assessment?" He supplemented his initial response of "training," with this observation: "It wasn't until we found an assessment where we could have immediate turn-back of information that we actually got that engagement with teachers and students." Kitchens also repeated his overarching theme of standardization of vocabulary. Common assessment vocabulary is essential, he said, to harnessing technology's potential.

Another participant asked for comparability studies of student performance on online vs. paper-and-pencil tests. She had read a study that showed that students performed less well on online tests than on paper-and-pencil tests. Those results surprised the panelists. To stress kids' adaptability, Olsen told how on the first day of TESA administration, the reading passages required students to not only scroll down but to scroll sideways across the page. They fixed the layout the next day. Even so, said Olsen, "no kid complained for four class periods." Ruess also pointed to *digital natives*[2] comfort with and preferences for using these technologies. John Ross, associate director of IAETE and moderator of the session, cited studies by Walter Haney and Michael Russell[3] that demonstrated that students who were taught writing skills using a computer performed best when assessed on a computer as opposed to paper and pencil.

---

[2] A term coined by author and game developer Marc Prensky, which refers to the recent generation being born into and familiar with digital tools, as opposed to the older generation referred to as "digital immigrants."

[3] M. Russell, & W. Haney. *Testing Writing on Computers: An Experiment Comparing Student Performance on Tests Conducted via Computer and via Paper-and-Pencil* (Education Policy Analysis Archives, 1997), 5(3). Retrieved November 1, 2004, from http://epaa.asu.edu/epaa/v5n3.html

Leadership

- **Jan Barth,** Executive Director, Office of Student Assessment Services, West Virginia Department of Education
- **Lisa Brady Gill,** Executive Director, Office of Education Policy, Texas Instruments
- **Shelley Loving Ryder,** Assistant Superintendent, Division of Assessment and Reporting, Virginia Department of Education
- **Suzanne Triplett,** Program Director, State Support and Outreach, State Services and Constituency Outreach, National Assessment of Educational Progress, National Center for Education Statistics, U.S. Department of Education

West Virginia to Pilot Online Writing Assessment

We believe that assessment should be used to inform instruction first and foremost, [but also to] promote school improvement and for calculation of accountability.

—*Jan Barth*

West Virginia is entering the world of online assessment with a writing assessment. The policymakers' forum preceded the state's pilot program, scheduled for October 2004. Says Jan Barth, the state plans to be online with writing assessments by 2005.

As executive director of the Office of Student Assessment Services for the West Virginia Department of Education, Barth directs and manages a variety of statewide assessments. She believes classroom assessment is the high-yield ticket to close the achievement gap. West Virginia hopes to eventually have the writing assessment in its accountability plan. "We really believe multiple measures are about different tests, not different items in a test," explains Barth. For now, the online writing tests are not a part of how schools will calculate adequate yearly progress (AYP) as required under NCLB. That, says Barth, has created a high comfort level.

West Virginia is developing the assessment with CTB McGraw Hill. The state is starting with two grades (7 and 10), a decision made so as not to shut down the entire educational computing infrastructure. The two-grade introduction will require testing of roughly 44,000 students (22,000 per grade level) within a two-week window. For this task, Barth identified several basic challenges: infrastructure, issues of access and equity, maintaining the integrity and validity of data, security, and funding.

Infrastructure. Expanding the program will require increasing bandwidth and computer access for 20,000 students within a reasonable testing window. Both needs are being addressed.

Access to data. Speaking to sentiments expressed throughout the day, Barth stated, "We agree that results have to be back immediately. That's a really big issue for our assessment and accountability, which is called WESTEST." Describing the instructional component of WESTEST, Barth said, "This is a statewide summative test. We give back holistic scores, but we also give back analytical scores which speak to each student's strengths and weaknesses." Results are organized by school, county, and state. The state strives to provide data in a usable form along with research-based responses. However, Barth emphasizes, the district must have the will and personnel to implement curricular improvements.

Accessibility. Barth said WESTEST required a good deal of retrofitting for special needs populations. From this point on the state's plans proactively address those needs.

Integrity and validity of data. West Virginia will run a comparability study between the electronically scored writing samples and the paper-and-pencil versions. The state has collected longitudinal data since 1985. To preserve the value of those records, the state will maintain its original five analytics as it expands from a four-point rubric to a six-point rubric. That expanded rubric will align to NAEP and the writing samples for the ACT and SAT—providing greater comparability value to the test. Training the scoring engine with the new rubric will require at least 600 papers per prompt. Score distributions will determine if the engine needs further training. They have also created a "crosswalk" to WESTEST Performance Levels and made the reports similar.

Security. The exam will exert some control over the desktop as a security measure. There will be no access to Web sites, and no hotkey, menus, or right-click mouse functions. To ensure reliability, there will also be a variety of writing prompts.

Funding. Funds set aside for the project will enable schools to purchase more bandwidth and equipment. Federal dollars to develop assessment have freed up some funds for these needs. Every office in the West Virginia Department of Education, says Barth, is contributing something to the bandwidth need. Additional costs to assess grades 7 and 10 will include CTB Writing Roadmap software; training on the testing engine; data entry by a CTB account manager; a pilot program of 1,700 students; development, printing, and shipping of the manual; and some administrative expenses.

Virginia Encourages Online Testing Adoption as an Option

What we've found over time is that as people have participated, they have found the great benefits of online testing. Other divisions have heard about the value and have been interested in participating. They have not felt like they are being forced to do so.

—Shelley Loving Ryder

Virginia currently gives its school divisions (the Commonwealth's term for districts) the option of offering high school end-of-course tests online in many curricular areas. Implementation has been phased in and will eventually include all end-of-year tests for all grades. The ambitious online testing program began as part of a larger technology initiative that assisted districts to prepare their infrastructures for online testing and to improve use of the Internet and Web-based resources. Shelley Loving Ryder, assistant superintendent for assessment and reporting at the Virginia Department of Education, described the history of the policy decisions behind the program. Proceeds from bond sales funded the program, and she believes that providing schools with infrastructure funds prior to any online testing initiatives was crucial to success.

"Virginia is very much a local-control state," said Ryder, "so while the monies were provided to school districts, nobody in the state department or the legislature told the localities how to use them." The money was intended for school divisions to improve and certify their infrastructures as ready for online testing. That process included a list of architectural guidelines, a checklist of technical capabilities, and a load testing system.

At the time of the policymakers' forum, most school divisions were using the online testing for at least part of their assessment program. The remaining divisions were planning to come on board by the fall. Ryder explained that the department made online testing entirely voluntary, and its use has gradually increased as educators have seen the "great benefits."

The Commonwealth's legislature, Ryder stressed, strongly supported the online testing initiative. "One of the reasons Virginia went to online testing," says Ryder, "was it needed a faster turnaround of scores, and our legislature believed that online testing was the answer to that." The initiative was managed jointly within the department of education by the technology and assessment divisions. In retrospect, Ryder says, it would have been better to integrate the project through the assessment division.

Because Virginia already had a paper-based, high-stakes testing program in place, it decided that, at least initially, its online end-of-course testing would mirror the paper version as closely as

possible. Virginia's DOE put out an RFP in 2000 asking vendors to propose solutions for online testing. After piloting the most promising proposals, the state chose one submitted by Pearson Education (known at the time as NCS). Referencing Olsen's advice, Ryder said, "We did have 'kids on us.'" State officials held focus groups with students to identify the desired test-taking tools in an online environment. For example, because the assessments rely on multiple-choice questions, students identified being able to cross out distracters as a valuable testing skill that should be supported in the new environment. In addition, in answering questions based on reading passages, students said they preferred the passage and the item to scroll separately so that any part of the passage could be viewed at the same time as the item. There are some differences between the paper and online test versions, Ryder noted. For example, paper-based tests have several items on a page; the online version has only one item on the screen.

Virginia's first, full online implementation occurred in the fall of 2001. It was a deliberate decision to phase in implementation, explained Ryder. At the time of the policymakers' forum, the program was well established in high schools, just beginning in middle schools, and scheduled for implementation in elementary schools in 2009. In addition to phasing in grade levels, they gradually added more tests. The program began with Algebra 1 and reading in the fall of 2001 and, with each administration, more tests have been added. Once online tests are available from the state, schools can phase in their own mix of paper-based and online tests. The state has run comparability studies and hopes to make them available online.

In Virginia, serving special populations remains a challenge. If she were to do it again, Ryder said, she would address accommodations for these students proactively. The state is now retrofitting its extensive list of accommodations and encountering surprising difficulties. Though it would seem that large print is a natural for the online environment, the state's graphics-based software cannot easily enlarge text. The read-aloud accommodation originally required a reader to stand behind the student at the computer. That proved to be too intrusive, so now both the reader and the student can view their own monitors.

As it looks ahead, Ryder says the Commonwealth hopes to take greater advantage of the new online medium and to look at how technology can deliver items in a different way. For now, paper and online versions will be the same. However, as Virginia creates new middle school assessments in reading and mathematics to meet the requirements of NCLB, it is, for the first time, field-testing items in both paper and online delivery modes. Ryder also looks forward to using testing IDs, which will support better tracking of scores over time.

Three Online Testing Initiatives at NAEP Define the Work Ahead

We ran into a lot of challenges, but . . . we think we still have to go there—and probably sooner rather than later.

—*Suzanne Triplett*

In roughly two weeks in February 2005, the U.S. Department of Education's National Assessment of Educational Progress (NAEP) (http://nces.ed.gov/nationsreportcard/) will administer tests to 1,250,000 students in about 20,000 schools across the country. The federal agency will hire 5,000 field staff to administer those tests and conduct 600 sessions a day. Suzanne Triplett, director of state services and constituency outreach for NAEP, explained that this massive testing program has had three forays into technology-based testing.

Triplett prefaced her comments with an explanation of how NAEP differs from most state assessments—and thus the instructional and classroom priorities of most attendees. "We are a national assessment. We report at the national, state, and district level. We do not, at least at this point, provide any school or student results. We're designed not to do that. We are prohibited by law from influencing instruction."

NAEP's Math Online (MOL) study, administered in 2001, simply tried to put existing paper-based items online. Writing Online (WOL), administered in 2002, compared responses from traditional paper-and-pencil writing assessments with computer-delivered prompts and responses. The Technology Rich Environment (TRE), the most recent effort, examined eighth-grade students' ability to explore and synthesize scientific information online. Similar to Harmon's efforts in Georgia, Triplett said, "We tried to use computers to test things we couldn't test using paper and pencil."

However, Triplett said, the agency "ran into some problems." Even with the simplicity of the MOL assessment, problems compromised the fidelity of the presentation, such as when rulers appeared differently with changes in computer settings. Triplett also noted that it was labor intensive to manage the infrastructure and intrusive to schools for NAEP to take over labs or instructional computers. NAEP had anticipated a lack of standardization of hardware within states, but the agency also discovered it within schools. Student differences were a complication as well. Among students, there was "enormous variation in skills, even within the same classroom," Triplett said. The tests are timed, which adds another layer of complexity.

Simply changing existing items to the technology-based format, Triplett explained, took a long time. It also took "a very long time" to develop new items. It was expensive, Triplett said, and she questioned the assumption that large, upfront development costs would be offset by savings down the road. Triplett also acknowledged the much larger issues involved, such as measurement, validity, and equity.

"We did find that students love it," said Triplett of the online testing. "They are much more engaged." She also thinks test creators must work to make their items more interesting, and said, "If it is just the same old thing, we don't think it is going to work." On the positive side, she noted, "This is a wonderful opportunity for us to test special needs populations more effectively than we ever have."

A move to technology-based assessment, in Triplett's opinion, is inevitable. "I don't think we have a choice," she said. "I think that's where this generation of kids is going. We can't measure what they know and can do, if we don't move in this direction." She then added, "If Virginia is way out here with online assessment and NAEP comes wandering in every few years with pencil and paper, we have a big issue."

Lobbying for Assessment Technologies

Form a crisp story around the educational benefit, even if you don't have the SBR [scientifically based research] yet. They really do want to hear those real-life stories and see those real-life examples.

—Lisa Brady Gill

Lisa Brady Gill, executive director of the Office of Education Policy in the Educational and Productivity Solutions Division of Texas Instruments (TI), explained that her company had multiple stakes in attending the symposium. First, TI has an educational technology division, and one of Gill's roles is to help customers implement education policy. Second, like all businesses, TI relies on an educated workforce. According to Gill, business has a substantial and growing influence on the definition of what students need to know to enter the work force. Businesses want to ensure that schools teach the twenty-first century skills of analyzing, collaborating, and teamwork, she said. When technology is purchased by schools, businesses and the community at large want proof that it is effective and integrated into instruction. They want to know it is a part of a well-designed curriculum used by highly trained teachers.

Gill said educators come to her all the time wondering whom they should talk to, given limitations on their ability to lobby. More specifically, these educators express a desire to explain the importance of formative, classroom-based assessment to policymakers. Gill advised the audience, as she has advised educators, to communicate with business organizations, discuss actionable steps, and take them.

Speaking from her company's perspective and experience with assessment technology, Gill said that from the late 1980s to the late 1990s, it became increasingly acceptable and important for students to use calculators on tests. When using the same tools on assessments that students use when learning and teachers use in instruction emerged as an accepted educational goal, the use of graphing calculators increasingly became a part of national and state standards. During that time, said Gill, educators controlled standards. In the past few years, however, many other groups, including business and federal policymakers, have gained more influence. She pointed to standards committees made up of experts from around the world and the influence of the business community in legislating NCLB.

This shift has heightened the need for educators to talk to policymakers and tell them stories that demonstrate the benefit of organizations that influence education policy. Gill said the Consortium for School Networking (CoSN), The Software and Information Industry Association (SIIA), The International Society for Technology in Education (ISTE), The American Electronics Association, and The Business RoundTable "all have active positions in educational technology and in education policy and they want to represent education's views."

She also encouraged stories of the educational benefits of school technology. "When we go on Capitol Hill, they tell us they don't hear those stories enough. . . . They don't really want to hear it from TI," said Gill. "They want to hear it from you." Gill added that policymakers need to be convinced of the return on investment. She closed by encouraging schools and educators to continue to work with their business partners to develop the tools that are needed. "We want to work with you to create the systems and the tools that you need to become successful in your states," she said.

Listening to Leadership's Experience

The participant discussion following the panel presentation revealed strong endorsement of open communications and the resulting actions but, said NAEP's Triplett, "I don't want you to

go away thinking those are easy things, because they are not." Those attendees who were moving technology-based assessment forward underscored the difficulties of their efforts. They said infrastructure is an enormous issue. Some said they have often put the cart before the horse. Many identified the error of retrofitting the possibilities to serve students with special needs instead of innovating from the beginning. Others pointed to standards that require assessments that have not yet been developed. Data cannot yet move easily between testing and reporting systems, and substantial professional development is needed if teachers are to make sense of the data. While these speakers could anticipate improvements with entirely new types of assessment items or new technologies, such as handheld units, that kind of change seems distant, given their struggles with acceptance of and the technical expertise required for their most basic projects. The leadership panel gave a strong dose of reality to the conversation.

Soapbox Live

- **John Lee,** Senior Researcher, Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA

- **Pat Roschewski,** Director of Statewide Assessment, Nebraska Department of Education

The afternoon portion of the symposium, moderated by John Ross, IAETE's associate director, brought ideas generated in an e-mail-based forum sponsored by IAETE at AEL to a live audience. Soapbox (www.iaete.org/soapbox) explores educational issues related to emerging technologies by gathering diverse groups of experts to participate in weeklong e-mail-based discussions. One Soapbox panel just prior to the policymakers' symposium praised the impressive instructional benefits of electronic student portfolio assessment and bemoaned the reliability and validity issues that make it difficult to use those artifacts for state accountability. At the symposium, two more experts extended the portfolio discussion to the issues involved in using technology for data management, sharing their own experiences within a policy context.

John Lee, a senior researcher at the Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA, helped develop the Quality School Portfolio (QSP), a free, technology-based portfolio system from CRESST that helps schools manage assessment data. Pat

Roschewski, director of statewide assessment for Nebraska's department of education, helped the state develop a system in which each district creates a portfolio of classroom assessments to meet state accountability needs.

QSP, A Tool for School Improvement

The whole purpose behind our project is using data to improve student performance. . . . Typically you have a lot of different data that is in a lot of different places, and you want to put it together in one place. That's what QSP allows you to do.

—John Lee

The QSP portfolio tool serves as a central repository for a variety of school assessment data, allowing them to be tracked over time. QSP generates valuable reports for district administrators, principals, teachers, students, and parents. It can help these groups measure progress toward standards and generate reports for accountability purposes. The tool is divided into five main sections:

1. **Groups:** Disaggregates student, teacher, and parent data into custom-designed "Groups" for analysis and reporting

2. **Goals:** Determines goals and sets targets to monitor student progress toward meeting standards

3. **Reports:** Creates understandable and actionable charts and graphs as a basis for making decisions

4. **Gradebook:** Tracks student performance at the classroom level

5. **Students:** Stores and organizes student work samples, providing a longitudinal history of each student. This section also houses the digital portfolio, which allows artifacts to be linked to descriptors and rubrics.

"It is important to bring in different types of data," explained Lee. "It gets at a much broader picture." The new Web-based version of QSP includes the ability to incorporate learning data, demographic data, perception data, and achievement data. As Lee explained, data do not arrive from test makers ready to run in QSP; however, the group that is continuing to build QSP is working closely with the Schools Interoperability Framework (SIF) to provide input on developing standards that would support this capability. At this time, data need to be cleaned, such as

being checked for errors in student identification numbers during batch entry. In the end, the data are stored in a relational database that allows users to query and interpret data with great flexibility.

The QSP project, which began nine years ago, has been funded primarily by the U.S. Department of Education. The Web-based version was rolled out about two years ago and was being used in 26 states at the time this forum was held. One hundred twenty districts across the country, with more than 900,000 students, were using it, many as members of consortia. In Michigan, for example, a consortium of about 60 districts, with 150,000 students, runs QSP from a single server. The original desktop version of QSP is being used in all 50 states.

The Web-based QSP has been implemented at various levels in districts across the country, thus providing CRESST with input for the continued development of the tool. As an example of recent additions, Lee cited the ability to include not just a letter grade for an assessment but also a proficiency rating for each relevant standard and a standards-based progress report.

Training on the use of QSP is available for a fee from partners across the country. QSP's online training covers both the use of the QSP software and the broader issues of data-based decision making. "The cycle of investigation," explained Lee, "is a very iterative process of questioning." He adds that it is important to take action along the way, rather than just running a lot of reports, because action is ultimately what will make the difference for students.

Nebraska Using Portfolios of Classroom Assessments for Accountability

Philosophically, every policy that was formed needed to be framed around two questions: (1) What's best for kids? and (2) How do we bring the level of professional development to the teachers so that they can . . . have confidence in the information they are getting from classroom assessments? How can they make their assessments of sufficient quality to be used for state purposes?

—Pat Roschewski

Before 2000, Nebraska had no legislation for state standards assessment or for an accountability system. When federal law required it, officials had to build a statewide system for use by districts that fiercely guarded local control. During this process, said Pat Roschewski, director of Statewide Assessment for the Nebraska Department of Education, the goal that remained first and foremost was "student learning and the achievement of our kids."

Nebraska modeled its state system on concepts that Roschewski, formerly a district administrator, originated in her district. Because of the difficulties capturing student learning, Roschewski's district contracted with the Buros Center for Testing (http://www.unl.edu/buros/) and posed the question, How do we become assessment literate? She never guessed that the system that evolved from this partnership would become a state—and possibly national—model. Roschewski's model was adopted by the state and is known as STARS, School-based Teacher-led Assessment Reporting System (http://www.nde.state.ne.us/stars).

STARS requires all school districts to adopt the state standards, or to submit their own standards of greater rigor for review. All districts are required to assess those standards in their local assessment systems. That local system includes a statewide norm-referenced test that assesses about 30 percent of the state's curriculum standards. Progress toward the remaining standards is measured with teacher-created classroom assessments that are integrated with instruction.

After giving the audience a chance to picture the system, Roschewski explained, "There is a set of very rigorous technical requirements that districts have to meet in order to be able to use those classroom-based assessments." Districts submit their assessment portfolios to the state. The state, through a contract with the Buros Center for Testing, contracts with assessment experts from across the country, who rate and provide feedback on the local assessment systems. In the final accountability report, districts are given two ratings: (1) on the quality of their local assessment and (2) on the performance of the students on that assessment.

Building the technical infrastructure to support this effort required four different systems. Initially, assessment plans were submitted to the state on paper. Just this year, Nebraska field-tested electronic submission in 30 districts with great success. They built another system to collect information on students and standards from every district. A third system supports the state's writing test, for which 800 trained scorers work at the local level to turn data around within two weeks. Finally, they built a system to warehouse, manipulate, and display test score data.

Demographics in Nebraska show that 300 of the state's 501 districts have fewer than 10 students. Eleven districts have no students. Eighteen have fewer students than school board members. As could be expected, these unique population numbers made meeting technical support needs difficult. Even so, the state department of education got the final three systems up and

running in less than eight months for the 2000-2001 school year. "They have embraced the electronic systems," said Roschewski. "Four years into the process, there are no more questions. Everybody is up to speed."

Together, these four systems annually generate a "massive report" known as *The State of the Schools Report*. Said Roschewski, "It comes out every fall. It really is the opportunity for anybody to drill down to standard-level information in any building or in any district in our state."

Nebraska explored packages to purchase, but "nothing fit," Roschewski said. "The University of Nebraska was our partner in two of the systems that we built, and literally built them for nothing." The university created an "assessment plan" submission system and the "District Assessment Portfolio" submission system. An Omaha-based branch of Quilogy, a private company, developed the input system for reporting on standards, the *State of the Schools Report* display of all the data, and the data collection/scoring system for the statewide writing assessment. All of the systems are Web-based. Of this experience and its success, Roschewski advised, "We cannot minimize the importance of the upfront discussion with the contractor. I found that those hours sitting with those folks around the table, discussing the whole thing, were critical hours in terms of the final output."

Much of the professional development is accomplished through intermediate agencies. "The price of keeping decisions at the local level has been significant in terms of time and in terms of resources," said Roschewski. However, the state is shifting money into professional development that would have gone to test makers, and Roschewski sees that as the preferred investment. "Teachers have always assessed," she said. "What we had to do was teach them how their assessment was of sufficient quality to make confident inferences from those data."

Putting It Together

Both the QSP and Nebraska assessment portfolios reflect the desire to define student work with a mix of tools—and both try to get as close to the classroom and to actual student work as possible. All three audiences for this series of discussions (education practitioners, researchers, and policymakers) share these goals and the hope that technology can accelerate progress toward them.

The 2002 symposium for education practitioners identified the urgency of having meaningful classroom assessments—formative assessments to shape instruction and intervention. Research-

ers at the 2003 event were looking forward to assessment focused on how student knowledge is organized, rather than on the recall of bits of information. Both groups expressed feeling that their priorities often seem to conflict with the new importance of using standardized tests used for state accountability. At the education practitioners' symposia, Chris Dede stated,

> We are in a "reform" movement, where powerful methods of teaching/learning are harder to use, due to flawed standards and tests. The only way to improve this situation is to give people something to move toward—not something to move against, because then we'll just react away from what we have now into some other flawed method of reform.

At the researcher's symposium, Dede expressed a similar desire to establish a goal for a new direction. Indeed, both groups looked toward the opportunities created by increased attention to assessment and the possibilities of new technologies. Speaking at the researchers symposium, Martin Orland, then special assistant to the director of the U.S. Department of Education's Institute of Education Sciences (IES) and acting director of the Office of Reform Assistance and Dissemination, addressed the administration's goals for "researched-based education" and its various implications for assessment. He stated, "We are not going to see in the next generation any improvement without getting assessments right." Speaking at the education practitioners' conference, Dr. Linda Roberts, a consultant who was previously the director of the U.S. Department of Education's Office of Educational Technology, observed, "First of all, the truth of the matter is, assessment is hot. The public's attention is on assessment and accountability. It is an incredible opportunity for us to improve what we do."

Those who must actually design policy for assessment are so deeply involved in pushing their way through these early days of new technology-based assessment that they have limited time for revolutionary changes. They are working to put infrastructure in place, to maintain data fidelity, and to meet a daunting list of legal requirements for inclusion, validity, and reliability. For this group, technology promises to capture, store, and manipulate data securely. Creating a new way of defining what students know is of interest, but it seems like a distant goal.

Together, the three IAETE-sponsored symposia give a comprehensive picture of the work and possibilities ahead for the development of technology-enhanced assessment. Teachers, students, and parents are becoming increasingly aware of assessment issues and individual assessment data. A growing understanding of the purpose, benefits, and limitations of various assess-

ments will enable new assessments to be added to the mix. Some assessments will be entirely new, such as concept maps or complex simulations or virtual environments. Others will be familiar but more rapidly and reliably scored with technology. Their impact on instruction will be limited until data from these types of assessments are valued, as demonstrated by inclusion in accountability measures. Even so, experience with their use and the revelation of their potential to improve instruction and student achievement are likely to create advocates.