Perspectives on the Integration of Technology and Assessment

James W. Pellegrino
University of Illinois at Chicago

Edys S. Quellmalz
WestEd

Abstract

This paper considers uses of technology in educational assessment from the perspective of innovation and support for teaching and learning. It examines assessment cases drawn from contexts that include large-scale testing programs as well as classroom-based programs, and attempts that have been made to harness the power of technology to provide rich, authentic tasks that elicit aspects of integrated knowledge, critical thinking, and problem solving. These aspects of cognition are seldom well addressed by traditional testing programs using paper and pencil or computer technologies. The paper also gives consideration to strategies for developing balanced, multilevel assessment systems that involve articulating relationships among curriculum-embedded, benchmark, and summative assessments that operate across classroom, district, state, national, and international levels. It discusses the multiple roles for technology in an assessment-based information system in light of the decision support needed from the multiple actors who operate across levels of the education system. The paper concludes with a consideration of the current state of the field as well as the potential for technology to help launch a new era of integrated, learning-centered assessment systems. *(Keywords: Assessment, technology, large-scale, classroom, formative, summative)*

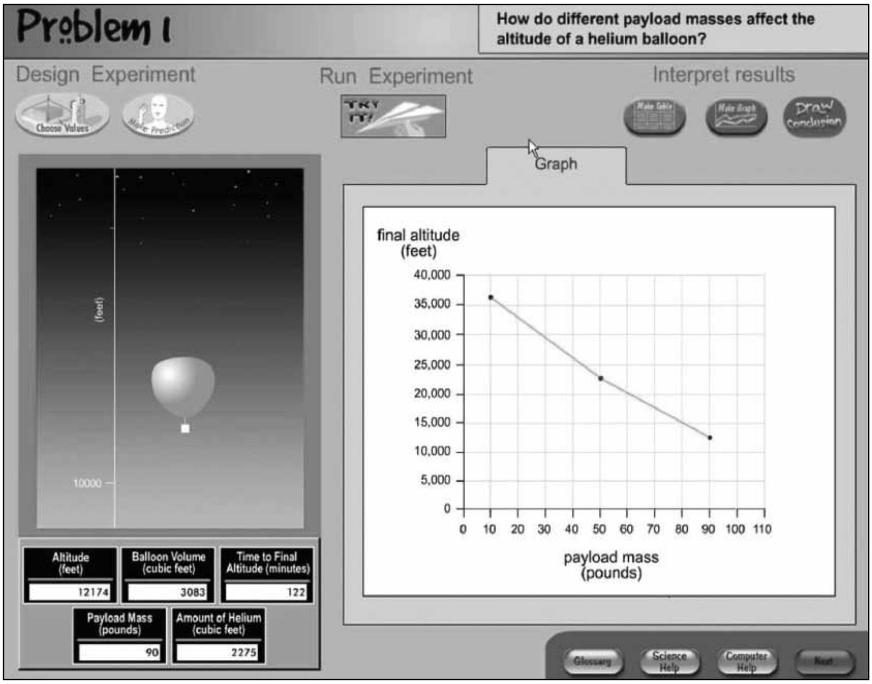
Across the disciplines, technologies have expanded the phenomena that can be investigated, the nature of argumentation, and the use of evidence. Technologies allow representations of domains, systems, models, data, and their manipulation in ways that previously were not possible. Dynamic models of ecosystems or molecular structures help scientists visualize and communicate complex interactions. Models of population density permit investigations of economic and social issues. This move from static to dynamic models has changed the nature of inquiry among professionals as well as the way that academic disciplines can be taught. Correspondingly, a new generation of assessments is well on its way to transforming what, how, when, where, and why assessment occurs and its linkages to teaching and learning. Powered by the ever-increasing capabilities of technology, these 21st century approaches to assessment expand the potential for tests to both probe and promote a broad spectrum of human learning, including the types of knowledge and competence advocated in various recent policy reports on education and the economy (e.g., NCEE, 2007; NRC, 2006).

Although early uses of technology in large-scale testing have focused on relatively straightforward logistical efficiencies and cost reductions (see e.g., Bennett, 2008; Quellmalz & Pellegrino, 2009), a new generation of innovative assessments is pushing the frontiers of measuring complex forms of learning. The computer's ability to capture student inputs permits collecting evidence of processes such as problem-solving sequences and strategy use as reflected by information selected, numbers of attempts, approximation to solutions, and time allocation. Such data can be combined with statistical and measurement algorithms to extract patterns associated with varying levels of expertise (e.g., Vendlinski & Stevens, 2002). Research in the learning sciences is simultaneously informing the design of innovative, dynamic, interactive assessment tasks and powerful scoring, reporting, and real-time feedback mechanisms. When coupled with technology, such knowledge has propelled various advances in adaptive testing, including knowledge and skills diagnosis, the provision of immediate feedback to teachers and students accompanied by scaffolding for improvement, and the potential for accommodations for special populations. Technology also supports movement toward the design of more balanced sets of coherent, nested assessments that operate across levels of educational systems.

Each of the preceding constitutes a major body of theory, research, and development and deserves major treatment that is well beyond the limits of the current article. This paper attempts to illustrate some of the major trends at work by examining a few of the emergent cases that have used technology to push the envelope with regard to a new generation of educational assessment. It seems clear that the use of assessment to support the attainment of many of our current goals for education improvement will require interdisciplinary partnerships and considerable additional research and development. It will also demand major shifts in education policies and practices in the designs of assessments, the models for testing, and the use of assessment data for various purposes, including student, teacher, and system-level accountability.

Technology-Enabled Assessments in State, National, and International Assessment Programs

Information and communications technologies such as Web browsers, word processors, editing, drawing, simulations, and multimedia programs support a variety of research, design, composition, and communication processes. These same tools can expand the cognitive skills that can be assessed, including the processes of planning, drafting, composing, and revising. For example, the National Assessment of Educational Progress (NAEP) writing assessment in 2011 will require use of word processing and editing tools to compose essays. In professional testing, architecture examinees use computer-assisted design



Problem 1

How do different payload masses affect the altitude of a helium balloon?

Altitude (feet)	Balloon Volume (cubic feet)	Time to Final Altitude (minutes)
12174	3083	122

Payload Mass (pounds)	Amount of Helium (cubic feet)
90	2275

Figure 1. Example of large-scale science assessment task developed by ETS.

(CAD) programs as part of their licensure assessment. The challenge that such technology-based presentation and data capture contexts offers now lies in the design principles for eliciting complex learning, the analysis of complex forms of data, and their meaningful interpretation relative to models of the underlying components of competence and expertise.

The area of science assessment is perhaps leading the way in exploring the presentation and interpretation of complex, multifaceted problem types and assessment approaches. In 2006, the Programme for International Student Assessment (PISA) pilot tested a Computer-Based Assessment of Science to test knowledge and inquiry processes not assessed in the paper-based booklets. The assessment included such student explorations as the genetic breeding of plants. At the state level, Minnesota has an online science test with tasks engaging students in simulated laboratory experiments or investigations of phenomena such as weather or the solar system.

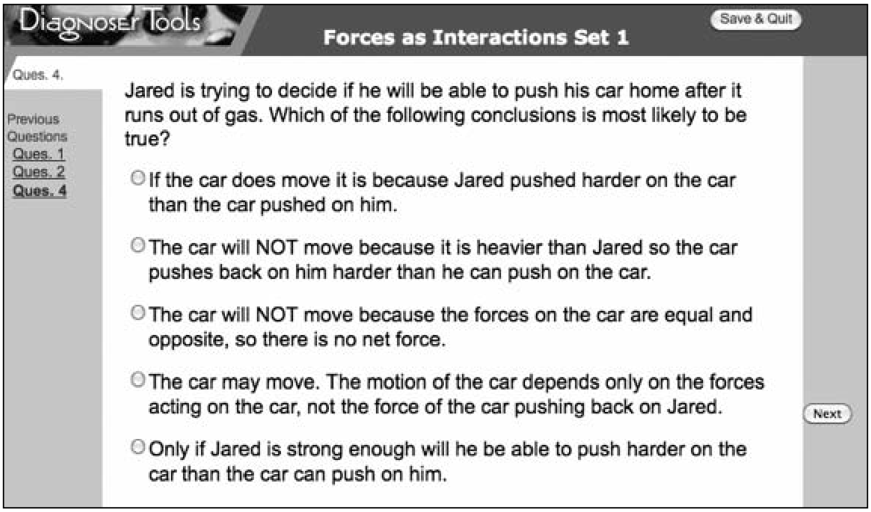
ETS pioneered the design of technology-based assessments for complex learning and performance (Bennett, Persky, Weiss, & Jenkins, 2007). An example of the type of item that Bennett et al. evaluated is shown in Figure 1. In this technology-based simulation task, eighth graders are asked to use a hot-air balloon simulation to design and conduct an experiment to determine the relationship between payload mass and balloon altitude. After completing the tutorial about the simulation tool interface, students select values for the independent variable payload mass. They can observe the balloon rise in the flight box and note changes in the values of the dependent variables of altitude, balloon volume, and time to final altitude. In another problem using the simulation, the amount of helium, another independent variable, is held constant to reduce the task's difficulty. Students can construct tables and graphs and draw conclusions by clicking on the buttons below the heading labeled Interpret Results. Figure 1 also shows the types of data that a student might obtain and plot prior to reaching a conclusion and writing a final response. As students work with the simulation, they can get help as needed in the form of (a) a glossary of science terms, (b) science help about the substance of the problem, and (c) computer help about the buttons and functions of the simulation interface that are built into the technology environment. Student performance can be used to derive measures of the student's computer skills, scientific inquiry exploration skills, and scientific inquiry synthesis skills within the context of physics.

The 2009 NAEP Science Framework and specifications drew upon ETS' work and other research in developing their rationale for the design and pilot testing of Interactive Computer Tasks (ICT) to test students' ability to engage in inquiry practices. Such innovative items were included in the 2009 NAEP science administration. The new 2014 Technology and Engineering Literacy Framework for NAEP will be entirely computer administered and will include specifications for interactive, simulation-based tasks involving problem solving, communication, and collaboration related to technology and society, design and systems, and information communications technology.

Large-scale testing programs such as those mentioned above are just beginning to explore the possibilities of using dynamic, interactive tasks for obtaining evidence of student content knowledge and reasoning. However, in the realm of high-stakes assessment for No Child Left Behind (NCLB) accountability, a number of regulatory, economic, and logistical issues have constrained the breadth and depth of the content and performance standards assessed in annual on-demand tests. Standard, multiple-choice item formats continue to dominate large-scale, computer-based, high-stakes testing, resulting in an overreliance on simple, highly structured problems that tap fact retrieval and the use of algorithmic solution procedures.

Technology-Enabled Assessments for Classroom Instructional Uses

A distinction has been made between assessments of the outcomes of learning, typically used for grading and accountability purposes (summative assessment), and assessments for learning, used to diagnose and modify the conditions of learning and instruction (formative assessment) (Stiggins, 2005). Research has repeatedly shown the formative use of assessment to significantly benefit student achievement (Black & Wiliam, 1998; Wiliam, 2007). Such effects depend on several classroom practice factors, including alignment of assessments with state standards, quality of the feedback provided to students, involvement of students in self-reflection and action, and teachers making



Forces as Interactions Set 1

Jared is trying to decide if he will be able to push his car home after it runs out of gas. Which of the following conclusions is most likely to be true?

○ If the car does move it is because Jared pushed harder on the car than the car pushed on him.

○ The car will NOT move because it is heavier than Jared so the car pushes back on him harder than he can push on the car.

○ The car will NOT move because the forces on the car are equal and opposite, so there is no net force.

○ The car may move. The motion of the car depends only on the forces acting on the car, not the force of the car pushing back on Jared.

○ Only if Jared is strong enough will he be able to push harder on the car than the car can push on him.

Figure 2. Example of a DIAGNOSER physics assessment item.

adjustments to their instruction based on the assessment results (Black et. al., 2004). Technologies are well suited to support many of the data-collection, complex analysis, and individualized feedback and scaffolding features needed for the formative use of assessment (Brown, Hinze, & Pellegrino, 2008). Two illustrative projects, one drawn from science and the other from mathematics, rely on detailed analyses of subject matter domains and student thinking to provide in-depth assessment and feedback during instruction.

The DIAGNOSER project is based on the facets framework for mapping aspects of student knowledge (Hunt & Minstrell, 1996; Minstrell & Stimpton, 1996) combined with principles of guided inquiry (see Minstrell & Kraus, 2005). It has set out to do the difficult work of breaking down physics concepts into requisite knowledge sets and misconceptions (facets). The facet framework is based on the understanding that students have preconceptions about scientific concepts that are not necessarily unique. For example, students may think that magnets exert a force only when they touch an object, or that “cold” can flow out of something that feels cold. For each of a number of physics concepts appropriate for middle school to high school level courses, researchers have created a series of multiresponse items designed to have every facet represented at least once in a response choice. Figure 2 provides an example of an item available online in the DIAGNOSER system. Occasionally, the system asks students to provide their reasoning for a response by choosing an option, which encourages consistent scientific reasoning. Each topic contains two question sets, and instructors are supposed to use prescriptive activities between question sets. Other resources are provided as support materials.

Based on a student’s response to an item such as that shown in Figure 2, as well as others in a set related to this topic area, the system diagnoses the student's level of understanding. Through the Web-based system, students respond and receive immediate and cumulative constructive feedback. Teachers receive the results immediately and can refer to a series of prescriptive activities tailored to address each facet. It would be difficult to replicate the feedback system without the Web-based design, which provides opportunities for self-regulated learning on the part of the student as well as targeted interventions on the part of the teacher. Providing two question sets along with supplemental material allows for intervention and re-assessment to work toward advancing student understanding. In a validation study in a Washington state district, students using DIAGNOSER outperformed their peers on items from the state science test.

Another example is the ASSISTment system, which is a pseudo-tutor for middle school level mathematics. Originally based on items from the Massachusetts state standardized test, researchers have developed a feedback system for each item through discussions with teachers. The system uses scaffolding questions, optional hints, and buggy messages (specific feedback given after student errors) for each item. Students must eventually reach the correct answer, and scaffolds/hints are limited to avoid giving away the answer (Feng, Heffernan, & Koedinger, 2006, 2009). Teachers receive feedback on student and class progress both on general summative measures (e.g., time to completion, percent correct) and on more specific knowledge components. Students also receive item-level analyses to identify specific issues with problems (Feng & Heffernen, 2005). Evaluation of the efficacy of ASSISTments has shown that performance is predictive of performance on randomly selected standardized test questions in paper-and-pencil format, and finer-grained models predict standardized-test performance better than typical scores (Feng et al., 2006, 2009), indicating that providing this analysis to teachers should be useful in interpreting students' skills. More than 60% of students self-report that the ASSISTments help them with the standardized tests, and there is some evidence that scaffolds help students transfer knowledge better than hints, especially on difficult problems.

In addition to assessment of student knowledge and skills in highly structured problems with one right answer, technology can also support the design of complex, interactive tasks that extend the range of knowledge, skills, and cognitive processes that can be assessed (Quellmalz & Haertel, 2004). For example, simulations can assess and promote understanding of complex systems by superimposing multiple representations and permitting manipulation of structures and patterns that otherwise might not be visible or even conceivable. Simulation-based assessments can probe basic foundational knowledge such as the functions of organisms in an ecosystem, and, more important, they can probe students' knowledge of how components of a system interact along with abilities to investigate the impacts of multiple variables changing at the same time (Quellmalz, Timms, & Buckley, 2009).

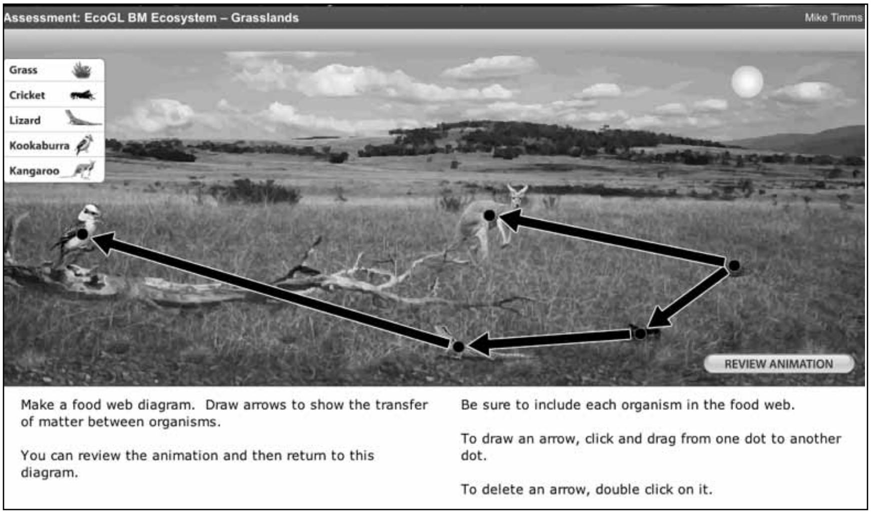
Moreover, because simulations use multiple modalities and representations, students with diverse learning styles and language backgrounds may have better opportunities to demonstrate their knowledge than are possible in text-laden print tests.

In an ongoing program of research and development, WestEd's SimScientists projects (http://simscientists.org) are studying the suitability of simulation-based science assessments as summative assessments with the technical quality required for components of an accountability system (Quellmalz, et al., 2008; Quellmalz, Timms, & Buckley, in press). New SimScientists projects are also studying use of simulations for curriculum-embedded formative uses of assessment. Figures 3 and 4 (pp. 126–127) present screenshots of tasks in a SimScientists summative, benchmark assessment designed to provide evidence of middle school students' understanding of ecosystems and inquiry practices after completing a regular curriculum unit on ecosystems.

Students are presented with the overarching problem of preparing a report to describe the ecology of an Australian grasslands ecosystem for an interpretive center. They investigate the roles and relationships of the animals, birds, insects, and grass by observing animations of the interactions among the organisms. Students draw a food web representing these interactions in the novel ecosystem. The assessment then presents sets of simulation-based tasks and items that focus on students' understanding of the emergent behaviors of the dynamic ecosystem by conducting investigations with the simulation to predict, observe, and explain what happens to population levels when numbers of particular organisms are varied. In a culminating task, students present their findings about the grasslands ecosystem.

In a companion set of curriculum embedded formative assessments situated in a different ecosystem, a mountain lake, the technological infrastructure identifies types of errors and follows up with feedback and graduated coaching. Figure 5 (p. 128) illustrates one of the levels of feedback and coaching that progresses from identifying that an error has occurred and asking the student to try again, to showing worked examples of investigations that met the specifications. For constructed responses, students self-assess by judging if their explanations meet criteria or match a sample response.

These SimScientists examples illustrate ways that assessment tasks can take advantage of the affordances of simulations to represent generalizable, progressively complex models of science systems; present significant, challenging inquiry tasks; provide individualized feedback; customize scaffolding; and promote self-assessment and metacognitive skills. Reports that the SimScientists learning management system generates in the embedded assessments for teachers and students indicate the level of additional help students may need and classify students into groups for tailored follow-on offline reflection activities, which further guide students to use scientific discourse. The project promotes model-based reasoning about the common



Make a food web diagram. Draw arrows to show the transfer of matter between organisms.

You can review the animation and then return to this diagram.

Be sure to include each organism in the food web.

To draw an arrow, click and drag from one dot to another dot.

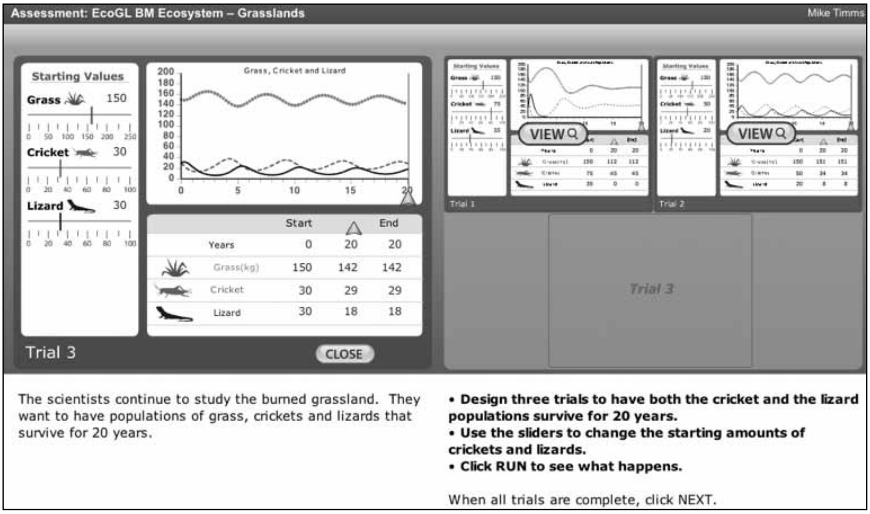
To delete an arrow, double click on it.

Figure 3. Screenshot of SimScientists ecosystems benchmark assessment showing a food web diagram interactively produced by a student after observing the behaviors of organisms in the simulated Australian grasslands environment.

organization and behaviors of all ecosystems to transfer knowledge about ecosystem components, interactions, and emergent behaviors to examples of new ecosystems. Research in the SimScientists projects is studying the technical quality of the assessments, the potential of the end-of-unit assessments as components of a state science accountability system, and the impact of the curriculum-embedded assessments and feedback on student learning. Project designs such as these can document the validity and utility of technology-based assessments for both instructional and accountability purposes.

Technology and the Development of Multilevel Assessment Systems

It is widely recognized that states must aim for balanced state assessment systems in which district, classroom, and state tests are aligned and mutually reinforcing. In the National Research Council (NRC) report *Knowing What Students Know*, a balanced assessment system relies on a nested system of assessments that exhibits features of comprehensiveness, coherence, and continuity (Pellegrino et al., 2001). Comprehensiveness is achieved by multiple measures of the full range of standards. Coherence involves a horizontal alignment of standards, goals, assessments, curriculum, and instruction as well as vertical alignment among assessments at different levels of the system. Continuity is achieved by going beyond annual, on-demand tests to multiple assessments over time and in time for teachers to tailor instruction. Indeed, underlying the introduction of the NCLB requirement for statewide testing is the assumption is that, given timely access to information about student learning, teachers will draw inferences about areas of need and use research and standards for best practice to make sound decisions for improving learning. However, the disconnect between classroom assessments and large-scale tests



	Start	△	End
Years	0	20	20
Grass(kg)	150	142	142
Cricket	30	29	29
Lizard	30	18	18

The scientists continue to study the bumed grassland. They want to have populations of grass, crickets and lizards that survive for 20 years.

- **Design three trials to have both the cricket and the lizard populations survive for 20 years.**
- **Use the sliders to change the starting amounts of crickets and lizards.**
- **Click RUN to see what happens.**

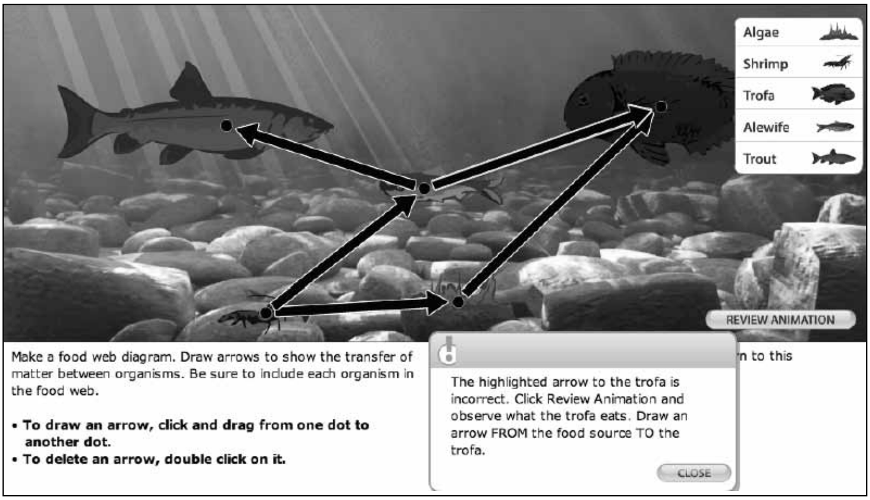
When all trials are complete, click NEXT.

Figure 4. Screenshot of SimScientists ecosystems benchmark assessment showing a student's investigations with the interactive population model.

tends to reduce the coherence of a statewide assessment system and interfere with sound instructional decision making.

Consensus is growing that better methods for capturing and connecting compelling evidence of student learning, both content knowledge and reasoning and inquiry skills, must be implemented across levels of the educational system. To this end, the National Science Foundation (NSF) funded the National Research Council to offer recommendations to states on their science assessment systems (see Wilson & Bertenthal, 2005). In a report commissioned by that project, Quellmalz and Moody (2004) proposed strategies for states to form collaboratives and use technology to create multilevel science assessment systems based on common standards and task design specifications. With the goal of helping schools and students meet federally mandated accountability goals, states are seeing formative assessment as a powerful tool for driving student achievement. Formative assessment is distinguished from summative, end-of-unit, and course benchmark assessments and from summative, interim, cross-unit assessments administered on a larger scale that are intended to describe the status of student performance periodically during the school year (Perie, Marion, & Gong, 2009; Quellmalz, Timms, & Buckley, in press).

When well designed and implemented, classroom assessments that are used during instruction to monitor and improve progress and that are also administered following instruction to document learning and identify remaining needs can become credible components of a multilevel state assessment system. Technology-enhanced formative assessments during instruction can provide immediate, contingent feedback and adaptive coaching for reteaching of problematic knowledge and skills. Benchmark assessments following instruction can provide summative classroom-based



Make a food web diagram. Draw arrows to show the transfer of matter between organisms. Be sure to include each organism in the food web.

- **To draw an arrow, click and drag from one dot to another dot.**
- **To delete an arrow, double click on it.**

The highlighted arrow to the trofa is incorrect. Click Review Animation and observe what the trofa eats. Draw an arrow FROM the food source TO the trofa.

n to this

Figure 5. Screenshot of SimScientists ecosystems formative embedded assessment providing feedback and coaching following student errors in drawing food web arrows after observing the behaviors of organisms in the simulated mountain lake environment.

assessments with technical quality that could be aggregated into the state accountability system.

Online authoring systems promise to streamline test design and reduce development costs (Mislevy & Haertel, 2007). The use of common specifications to design assessments at classroom, district, state, and national levels holds great promise for articulating the components of balanced assessment systems (Quellmalz & Moody, 2004). In addition, online design systems can support adaptations of assessments to offer accommodations for special populations while preserving the links between targeted standards and the designs of the tasks for eliciting evidence of achievement.

Technology's Multiple Roles in the Development of Assessment-Based Information Systems

Assessment programs should be designed to produce results that allow educators and policy makers to address a variety of questions about how a nation, state, district, school, program, group, or individual is performing. The goal is to make assessment results available to the right people, in the right form, at the right time so that assessment informs decisions and actions. Before individuals can answer questions, they must have the assessment results to work with. The goal is to collect the results in timely and efficient ways that allow educators and policy makers to answer their most important questions. Finally, it is not enough to answer questions; the answers must inform action. The goal is to help people use assessment results to make decisions about what actions to take. A technology infrastructure is a collection of tools and processes that help people achieve these goals.

Table 1: Functions of Technology in Supporting the Use of Assessment in Education

Collecting	Reporting	Using
Create assessment materials	Manage assessment data	Identify teaching resources
Tailor assessment administration	Analyze and interpret assessment data	Identify possible actions
Collect assessment data	Create reports	Design professional development
Score assessment performance	Distribute reports	Deliver professional development

There are a variety of roles that technology might play in this complex process. It is likely that a collection of tools, rather than a single tool, will fulfill these roles. However, information must flow appropriately between tools to support the differing needs of different people. Table 1 lists some of the roles that technology might play in the assessment process. Part of designing and implementing an assessment system for a school, district, or state is determining the appropriate roles for technology in that system.

Given the range of assessment tasks and data on student performance that are potentially available within the educational context, it is clear that the processes of collecting, scoring, and interpreting assessment data and then longitudinally tracking student performance are formidable. It is highly unlikely that any teacher, principal, or district or state administrator would be able to succeed in using all the assessment data available to him or her without the support of technology tools to assist in such a process across a variety of assessment levels and timeframes. Thus, an important direction for development and implementation of assessment systems is the design of specific and general technology-based tools that can assist educators in managing the assessment process. Ideally, one would like to have a system with extensive diagnostic assessment capability.

The design and deployment of even simple technology tools must ultimately rely on a technology infrastructure that connects the classroom to powerful database management and information retrieval systems that operate within and across schools and systems. This is especially true when the classroom assessment data are viewed as part of a coordinated system of assessment data that would potentially include curriculum-embedded assessment information, unit and end-of-course benchmark assessment data, interim cross-unit summative status checks, and state-level test data. Further work addressing issues of technology and the design of a comprehensive assessment system involves consideration of information and how it needs to flow through this system. For example, who needs to use assessment data? What questions need to be answered? In what timeframe do they need to be answered? What actions might they take based on these answers?

Conclusions, Implications, & Future Directions

It is an exciting time in the field of assessment for several reasons. First, individuals have realized that there are multiple roles for assessment to play in the educational process and that one of the most valuable roles is the formative function of assisting student learning. Second, cognitive research and theory have provided us with rich models and representations of how students understand many of the key principles in the curriculum, how students develop knowledge structures, and how to analyze and understand simple and complex aspects of student performance. Third, technology makes possible more flexible, tailored presentations to students of a much wider and richer array of tasks and environments where students can learn and where they can show us what they know and how they know it. Thus, there is an interesting and powerful confluence among theory, research, technology, and practice, especially when it comes to the integration of curriculum, instruction, and assessment.

In numerous areas of the curriculum, information technologies are changing what is taught, when and how it is taught, and what students are expected to be able to do to demonstrate their knowledge and skill. These changes in turn are stimulating people to rethink what is assessed, how that information is obtained, and how it is fed back into the educational process in a productive and timely way. This situation creates opportunities to center curriculum, instruction, and assessment around cognitive principles. With technology, assessment can become richer, timelier, and more seamlessly interwoven with multiple aspects of curriculum and instruction. As discussed earlier, the most useful kinds of assessment for enhancing student learning emphasize knowledge integration and extended reasoning, support a process of individualized instruction, allow for student interaction, collect rich diagnostic data, and provide timely feedback. The demands and complexity of these types of assessment can be quite substantial, but technology makes them feasible. Diagnostic assessments of individuals' learning, for example, must involve collecting, interpreting, and reporting significant amounts of information. No educator, whether a classroom teacher or other user of assessment data, could realistically be expected to handle the information flow, analysis demands, and decision-making burdens involved without technological support. Thus, technology removes some of the constraints that previously made high-quality formative assessment of complex performances difficult or impractical for a classroom teacher. The examples described above illustrate how technology can help infuse ongoing formative assessment into the learning process.

Clearly, we are just beginning to see how to harness technology to support the formative and summative functions of assessment. We still need to learn a great deal about the quality and efficacy of systems operating at both the large-scale level and the small-scale level. Not the least of the concerns facing us is the integration of assessment tools and practices into the educational system and teachers' practices. But we must also take note of the fact that extremely powerful information technologies are becoming as ubiquitous in educational settings as they are in other aspects of people's daily lives. Technologies are almost certain to continue to provoke fundamental changes in learning environments at all levels of the education system. Many of the implications of technology are beyond people's speculative capacity. Little more than 15 years ago, for example, few could have predicted the sweeping effects of the Internet and social networking on education and other segments of society. The range of computational devices and their applications is expanding exponentially, fundamentally changing how people think about communication, collaboration, problem solving, connectivity, information systems, educational practices, and the role of technology in society.

Although it is always risky to try to predict the future, it appears clear that advances in technology will continue to impact the world of education in powerful and provocative ways.[1] Many technology-driven advances in the design of learning environments, which include the integration of assessment with instruction, will continue to emerge and will reshape the terrain of what is both possible and desirable in education. Advances in curriculum, instruction, assessment, and technology are likely to continue to move educational practice toward more individualized and mastery-oriented approaches to learning, yet at the same time intertwine networking with resources, experts, and peers in problems requiring more complex forms of reasoning, problem solving, and collaboration. This evolution will occur across the K–16+ spectrum. To manage learning and instruction effectively, people will want and need to know considerably more about what has been mastered, at what level, by whom, with what levels of scaffolding.

Consider the possibilities that might arise if we integrate assessment into instruction in multiple curricular areas and collect the resultant information about student accomplishment and understanding with the aid of technology. In such a world, programs of on-demand external assessment, such as state achievement tests, might not be necessary. Instead, it might be possible to extract the information needed for summative and program evaluation purposes from data about student performance continuously available both in and out of the school context.

Technology could offer ways of creating, over time, a complex stream of data about how students think and reason, independently and collaboratively, while engaged in important learning activities. We could extract information for assessment purposes from this stream and use it to serve both classroom and external assessment needs, including providing customized feedback to students for reflection about their knowledge and skills, learning strategies, and habits. To realize this vision, additional research on the problem and data representations and analysis methods best suited for different learning goals, audiences, and different assessment objectives would clearly be needed—and is certainly doable.

We can therefore imagine a future in which the audit function of assessments external to the classroom would be significantly reduced or

---

[1] This scenario is adapted from one originally developed in Pellegrino et al. (2001).

even unnecessary because the information needed to assess students, at the levels of description appropriate for various monitoring purposes, could be mined from the data streams generated by students in and out of their classrooms. A metaphor for such a radical shift in how one "does the business of educational assessment" exists in the world of retail outlets, ranging from small businesses to supermarkets to department stores. No longer do these businesses have to close down once or twice a year to take inventory of their stock. Rather, with the advent of automated checkouts and barcodes for all items, these enterprises have access to a continuous stream of information that can be used to monitor inventory and the flow of items. Not only can business continue without interruption, but the information obtained is far richer, enabling stores to monitor trends and aggregate the data into various kinds of summaries. Similarly, with new assessment technologies, schools would no longer have to interrupt the normal instructional process at various times during the year to administer external tests to students, nor would they have to spend significant amounts of time preparing for specific external tests peripheral to the ongoing activities of teaching and learning.

Extensive technology-based systems that link curriculum, instruction, and assessment at the classroom level might enable a shift from today's assessment systems, which use different kinds of assessments for different purposes, to a balanced design that would ensure the three critical features of comprehensiveness, coherence, and continuity. In such a design, assessments would provide a variety of evidence to support educational decision-making (comprehensiveness). The information provided at differing levels of responsibility and action would be linked back to the same underlying conceptual model of student learning (coherence) and would provide indications of student growth over time (continuity).

Clearly, technological advances will allow for the attainment of many of the goals that educators, researchers, policymakers, teachers, and parents have envisioned for assessment as a viable source of information for educational improvement. When we implement powerful technology-based systems in classrooms, rich sources of information about intellectually significant student learning will be continuously available across wide segments of the curriculum and for individual learners over extended periods of time. This is exactly the kind of information we now lack, making it difficult to use assessment to truly support learning. The major issue is not whether this type of innovative assessment design, data collection, and information analysis is feasible in the future. Rather, the issue is how the world of education anticipates and embraces this possibility, and how it will explore the resulting options for effectively using assessment information to meet the multiple purposes served by current assessments and, most important, to enhance student learning.

Author Notes

James W. Pellegrino is a liberal arts and sciences distinguished professor and distinguished professor of education at the University of Illinois at Chicago, where he also co-directs the Learning Sciences Research Institute. Correspondence regarding this article should be addressed to James Pellegrino, LSRI, UIC, 1007 West Harrison Street, M/C 057, Chicago, IL 60607, E-mail: pellegjw@UIC.EDU

Edys S. Quellmalz is director of technology-enhanced assessment and learning systems in the Mathematics, Science, and Technology Program at WestEd. Correspondence regarding this article should be addressed to Edys Quellmalz, WestEd, 400 Seaport Court, Suite 222, Redwood City, CA 94063. E-mail: equellm@wested.org

References

Bennett, R. E. (2008). *Technology for large-scale assessment.* ETS Report No. RM-08-10. Princeton, NJ: Educational Testing Service.

Bennett, R. E., Persky, H., Weiss, A. R., & Jenkins, F. (2007). *Problem solving in technology-rich environments: A report from the NAEP Technology-Based Assessment Project (NCES 2007-466).* Washington, DC: National Center for Education Statistics, US Department of Education. Retrieved on June 1, 2010, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007466

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment.* London, UK: King's College.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 8–21.

Brown, J., Hinze, S., & Pellegrino, J. W. (2008). Technology and formative assessment. In T. Good (Ed.), *21st Century Education. Vol 2. Technology* (pp. 245–255). Thousand Oaks, CA: Sage.

Feng, M., & Heffernan, N. T. (2005). Informing teachers live about student learning: Reporting in the assistment system. *Technology, Instruction, Cognition and Learning*, 3, 1–14.

Feng, M., Heffernan, N. T., & Koedinger, K. (2006). Predicting state test scores better with intelligent tutoring systems: Developing metrics to measure assistance required. In M. Ikeda, K. Ashley, & T-W. Chan (Eds.), *Proceedings of the Eighth International Conference on Intelligent Tutoring Systems* (pp. 31–40). Springer-Verlag: Berlin.

Feng, M., Heffernan, N. T., & Koedinger, K. R. (2009). Addressing the assessment challenge in an online system that tutors as it assesses. User Modeling and User-Adapted Interaction: *The Journal of Personalization Research (UMUAI), 19*(3), 243–266.

Hunt, E., & Minstrell, J. (1996). Effective instruction in science and mathematics: Psychological principles and social constraints. I*ssues in Education: Contributions from Educational Psychology, 2*, 123–162.

Minstrell, M., & Krause, P. (2005). Guided inquiry in the classroom. In J. Bransford & S. Donovan (Eds.), *How students learn history, mathematics, and science in the classroom* (pp. 475–515). Washington, DC: National Academy Press.

Minstrell, J., & Stimpton, V. (1996). A classroom environment for learning: Guiding students' reconstruction of understanding and reasoning. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 175–202). Mahwah, NJ: Erlbaum.

Mislevy, R., & Haertel, G. D. (2007). Implications of evidence-centered designs for educational testing. *Educational Measurement: Issues and Practices, 25*(4), 6–20.

National Center on Education and the Economy (NCEE). (2007). *Tough choices or tough times: Report of the new commission on the skills of the American workforce.* Washington, DC: Jossey Bass.

National Research Council (NRC). (2006). *Rising above the gathering storm: Energizing and employing America for a brighter economic future.* Committee on Prospering in the Global

Economy in the 21st Century: An Agenda for Science and Technology. Committee on Science, Engineering, and Public Policy. Washington, DC: National Academy Press.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.) (2001). *Knowing what students know: The science and design of educational assessment.* National Research Council's Committee on the Foundations of Assessment. Washington, DC: National Academy Press.

Perie, M., Marion, S., & Gong, B. (2009). Moving towards a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice, 28*(3), 5–13.

Quellmalz, E. S., & Haertel, G. (2004). *Technology supports for state science assessment systems.* Paper commissioned by the National Research Council Committee on Test Design for K–12 Science Achievement.

Quellmalz, E. S., & Moody, M. (2004). *Models for multi-level state science assessment systems.* Report commissioned by the National Research Council Committee on Test Design for K–12 Science Achievement.

Quellmalz, E. S., & Pellegrino, J. W. (2009). *Technology and testing. Science, 323*, 75–79.

Quellmalz, E. S., Timms, M., & Buckley, B. (in press). Science simulations for assessment. *International Journal of Learning Technology.*

Quellmalz, E. S., Timms, M., & Buckley, B. (2009). *Using science simulations to support powerful formative assessments of complex science learning.* Paper presented at the annual meeting of the American Educational Research Association. San Diego, CA.

Quellmalz, E. S., DeBarger, A. H., Haertel, G., Schank, P., Buckley, B., Gobert, J., Horwitz, P., & Ayala, C. (2008). Exploring the role of technology-based simulations in science assessment: The Calipers Project. In J. Coffey, R. Douglas, & C. Stearns (Eds.), *Science assessment: Research and practical approaches* (pp. 191–202). WDC: National Science Teachers Association.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan, 85*(4), 324–328.

Vendlinski, T., & Stevens, R. (2002). Assessing student problem-solving skills with complex computer-based tasks. *Journal of Technology, Learning, and Assessment, 1*(3). Retrieved on June 1, 2010, from http://www.jtla.org; http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1010&context=jtla

Wiliam, D. (2007). Keeping learning on track. In F. K. Lester Jr. (Ed.), *Second handbook of mathematics teaching and learning* (pp. 1051–1098). Greenwich, CT: Information Age Publishing.

Wilson, M. R., & Bertenthal, M. W. (Eds.) (2005). *Systems for state science assessment.* Washington, DC: National Academy Press.