Technology Tools for Collecting, Managing, and Using Assessment Data to Inform Instruction and Improve Achievement

GENEVA D. HAERTEL, BARBARA MEANS, AND WILLIAM PENUEL

The purpose of this chapter is to examine the ways in which technology is transforming practices of assessment and educational decision making. We will make the argument that the use of information and computer technology for these purposes is becoming increasingly common, but that the various technology applications are still fragmented and education has yet to realize the full potential of using technology to integrate instructional activities, assessment, and data-informed decision making. The specific categories of technology-supported tools reviewed here are (1) student data management systems; (2) technology-supported assessments for accountability; (3) technology-supported formative assessments; and (4) classroom communication systems. We will first provide a brief sketch of the national policy context for educational technology, and then describe each of these categories of technology applications and highlight findings from available research on their effects on student achievement, concluding with a framework for their integration.

The Broader Educational Technology Policy Context

Since the early 1980s, educational settings have been irrevocably changed by the implementation of technologies in classrooms. We have witnessed the introduction of both the technical infrastructure of computers, wiring, software applications, laser discs, and digital cameras in classrooms nationwide and the implementation of the organizational infrastructure of technical support for all this equipment, including the

---

Geneva D. Haertel is Director of Assessment Design and Research at the Center for Technology in Learning at SRI. Barbara Means is Director of the Center for Technology in Learning at SRI. William Penuel is Director of Evaluation Research at the Center for Technology in Learning at SRI.

professional development of teachers and school administrators. Supporting the infusion of these technologies into our schools is a loosely knit coalition of policymakers, corporate executives, technology vendors, public officials, educators, and parents.

Early federally funded educational technology programs promoted access to instructional content for underserved students through telecommunications (Star Schools) and schools' involvement in developing innovative educational programs supported by technology (Technology Innovation Challenge Grants). Both of these programs were competitive, requiring Local Education Agencies and their partners to develop ideas for technology use that would compete against other grant proposals. Hence, federal support for K-12 technology integration was selective—more suited to catalyzing the development of model programs than to ensuring universal access.

In 1996, Congress created the Technology Literacy Challenge Fund (TLCF), the first program of block grants for the purpose of supporting the integration of technology into the K-12 curriculum. Over a period of 4 years, this legislation made nearly $2 billion available to states to build the technology infrastructure necessary for schools to move into the digital age. States received block grants under TLCF in amounts proportional to the number of low-income students served in their schools. Within states, districts competed for TLCF funds, writing proposals for how they would use the money. The national educational technology plan issued in conjunction with TLCF described four "pillars" or goals that the funds would support. The first two of these—making computers accessible to every student and connecting classrooms to each other and to the world outside of schools—were access goals. The other two goals were integrating "engaging" educational software into the curriculum and preparing teachers to use and teach with technology.

Technology access was also the focus of the "E-rate," a universal discount phone service subsidy providing assistance in the form of subsidized Internet connection rates and payment for the cost of wiring classrooms in schools. The program was intended to help improve access to educational technology, especially in rural areas and in low-income districts. At the same time, private ventures such as Net Day enlisted massive numbers of volunteers to help wire schools for Internet access.

Statistics collected by the National Center for Education Statistics (NCES) documented remarkable progress in providing access to computers and the Internet within K-12 schools during the 1990s. The number of public school classrooms with Internet access rose from 3% to 87% between 1994 and 2001 (NCES, 2005). This progress, coupled with the achievement and accountability focus of No Child Left Behind (NCLB), heralded a new emphasis in national educational technology policy. John Bailey, then director of the U.S. Department of Education's Office of Educational Technology, announced "We have reached an important technology goal by connecting our schools to the Internet. . . . Now we must use these connections for a far more important goal of improving student achievement" (U.S. Department of Education, 2002).

Federal support for integrating technology into schools was authorized under Title II of the NCLB Act and reflected that legislation's emphasis on raising achievement and using scientific evidence as the basis for selecting educational practices. According to this legislation (Public Law 107–110; Part D, Section 2402), a key purpose of the program was to provide assistance to states and localities for the implementation and support of a comprehensive system that effectively uses technology in elementary and secondary schools to improve student academic achievement.

Federal grants to states to support K-12 educational technology were continued in the form of the Enhancing Education Through Technology (EETT) program. EETT called on states to distribute EETT funds to districts serving low-income students, with half the funds distributed on a competitive basis (as was done under TLCF) and half distributed by a formula based on the number of low-income students that the district served. During the first year of EETT (fiscal year 2002), the funding level of $700 million was a high point for educational technology grants to states, but subsequently funding levels were cut, and by FY 2006 were down to about $250 million a year. This funding reduction for educational technology no doubt reflected general pressures on the federal budget, but also was prompted by the assumption that the challenge of access had been largely overcome and by uncertainties about the value of many of the ways that technology was being used in schools.

A new national educational technology plan was issued in 2004, and that plan included a goal that had not appeared in earlier national plans—to use integrated and interoperable data systems to allow teachers and administrators to access student, subgroup, classroom, and school-level data for the purpose of raising student achievement. This use of technology is the first category of applications to be reviewed in this chapter.

Student Data Management Systems

The 2004 national educational technology plan (U.S. Department of Education, Office of Educational Technology, 2004) emphasized the use of interoperable data systems to improve education. The plan articulated the expectation that such data systems would provide multiple benefits to classrooms, schools, and districts, including increased efficiency in the operation of districts and schools, better alignment of resources meant to be allocated to those students with the greatest achievement needs, streamlined purchasing, and more frequent and appropriate use of assessment results to inform and differentiate instruction for every child.

As the plan notes, most school districts were far from the idealized goal of integrated, interoperable systems. The average school district worked with multiple "silos" of data concerning different functions, from food service to buses to instructional programming and planning to student achievement to purchasing and resource allocation. Lack of integration and interoperability of data systems prevented all parties from seeing the "big picture" to efficiently coordinate data to drive policy and decision making. The plan touted the ability of an interconnected and interoperable network from which data could be mined to highlight relationships between different system components, resulting in greater efficiency, better allocation of resources, and better alignment between instructional practice, professional development, and assessment. Most importantly, the use of data from such systems to improve instruction for students was expected to enhance achievement. The plan cited Poway Unified School District, a suburban district near San Diego, California, as an example of a district using data to enhance student performance. Poway Unified had built a data warehouse that integrated data from the district's student information, human resources, special education, and assessment systems and made it available to teachers. Teachers were expected to use the database to develop queries about their students' performance and then use the data to differentiate instruction based on those results.

Over the past 4 years, the assessment and data reporting requirements of NCLB have motivated states and districts to adapt or acquire database systems capable of generating the required student data reports. A national Data Quality Campaign has articulated 10 essential features for student data systems, notably the use of a unique student identifier that enables longitudinal tracking of a student's progress from year to year (see http://www.dataqualitycampaign.org/ for details). In addition to serving accountability reporting requirements, these student data systems are promoted as the foundation for data-driven decision making. The use of data to inform decision making has long been central to business concepts of "total quality management" (Deming, 1986), and the idea that education would benefit from a similar focus on collecting and reflecting upon data has been promoted for several decades (Popham, 1995; Schmoker & Wilson, 1995). But the increase in the amount of annual testing conducted by states and the requirements for analysis of student subgroup performance and performance changes over time, coupled with the consequences of the failure to make progress that are part of NCLB, gave a huge impetus to the data-driven decision-making movement. As one journalist summarized, "There is no denying that an integral part of the business of K-12 education today is to collect, manage, analyze, and learn from a wide array of data" (Salpeter, 2004). The hierarchical nature of school systems and the fact that student data are generated at the classroom, school, and district levels and used both at these three levels and by state and federal decision makers call for technology tools and interoperable data systems. Since the passage of NCLB we have seen an explosion of student data systems, system use guidelines, and professional development offerings coming out of both the public and the commercial sectors (see Wayman, Stringfield, and Yakimowski, 2004 for a review of many of these commercial software products).

Many of these commercial products involve multiple modules so that districts have the option of purchasing one or all of them. Typically systems include student assessment data, student demographic data and educational history (e.g., English language learners [ELL] status, attendance), teacher professional and personnel data, perceptions data about school climate from relevant stakeholders, and school program data (e.g., effectiveness of curricular programs). Wayman (2005) loosely categorizes data systems into three types: (1) student information systems that provide real-time accounting of student daily school functions (e.g., attendance, disciplinary actions), but are not available beyond the current school year; (2) student data from standardized, large-scale assessments that readily organize and analyze frequent benchmark assessments but do not provide such data over time; and (3) data warehousing systems that provide historical data of all kinds but are not available for immediate turnaround of analyses of new data.

Although many of these systems can accommodate any type of student data that schools and districts choose to incorporate, in practice, the databases typically contain little in the way of student learning outcomes other than scores on mandated tests. Significantly, there are few data on instructional practices or program participation that would allow district officials, local school leaders, or teachers to link learning outcomes to specific instructional strategies. In a similar vein, many of the systems support flexible data queries on the part of users, but most usage involves the generation of standardized reports.

Empirical Findings

Research on the use of student data systems has burgeoned during the past 5 years. Several researchers have provided positive reports of successful implementations (Cromey, 2000; Feldman & Tung, 2001; Light, Wexler, & Heinze, 2004). The implementation of the Grow Network reporting system in New York City was studied by Light and colleagues using data from 15 city schools. Surveyed and interviewed teachers reported using assessment data for their students in both broad and specific planning to help set whole-class priorities, make weekly and yearly lesson plans, construct specific daily lesson plans and mini-lessons, individualize instruction, group students, and pair high- and low-achieving students for peer tutoring. Student achievement was not measured using assessment data; thus, it is unclear how much of the information obtained from the Grow Network system simply confirmed what teachers already knew about their students' achievement levels.

Wayman and Stringfield (2006) examined the efforts of three schools to involve entire faculties in a review of student data that was supported by efficient data systems. The researchers reported results that indicated that administrative supports were key in fostering the use of such data systems and that improved teaching practices were often associated with the use of data. However, they were circumspect in their conclusions about the impact of such systems on student achievement, saying

> Finally, our study was not designed to address the causal relationships among data use, educational practice, and student learning. The teachers in our study felt the improvement in their practice benefited students and led to improved student learning, and administrators cited increased test scores. Still, without a longitudinal experiment, it is impossible to extend beyond these anecdotal associations, nor is it possible to attribute test score increases to this or any other initiative. Future research should be conducted experimentally that demonstrates the effects of data initiatives, such as those undertaken in these districts, on student performances and learning. (p. 569)

Herman and Gribbons (2001) studied a high school where overall student performance looked good when compared with national norms, but that had a subgroup of economically disadvantaged ELLs from outside the local area who had poor test scores. Looking at the data more closely, the staff discovered that poor attendance and lower enrollment in more demanding mathematics courses were associated not so much with where students lived as with where they had attended middle school. A misalignment between the mathematics sequences at the high school and some of the middle schools outside the local area made the transition into higher-level math courses in ninth grade difficult for these students. This discovery led staff to the understanding that ELLs' math deficiencies were based more on prior learning opportunities than on socioeconomic status and helped the faculty identify the kinds of learning opportunities that students from these particular middle schools would need to successfully transition into the higher-level mathematics courses.

One conclusion that can be drawn from several of these case studies is that schools that implement data-driven decision making successfully have a preexisting common vision for the district as well as a supportive leader who promoted the data management tool and set aside time for teachers to engage in data review and reflection (Cromey, 2000; Herman & Gribbons, 2001). However, this optimistic picture is balanced in the literature by a number of concerns. Many teachers and administrators express distrust of the standardized test scores that these systems typically employ (Cromey; Herman & Gribbons). While standardized-test data can be viewed as useful in framing annual analyses of student progress, such scores are inadequate for formative and diagnostic assessment (Thorn, 2002). Teachers express a preference for "real-time" data relevant to their professional circumstance (Mason, 2002).

Even promoters of data-driven decision making acknowledge that training of teachers and administrators in the use of data to make decisions is an area of need (Herman & Gribbons, 2001). Standard reports generated by assessment systems are often confusing to teachers and it is not clear that they appreciate such basics of assessment as "measurement error" and "sampling bias." Research has also shown that teachers and administrators using assessment data tend to focus on "bubble kids" just below the criterion for proficiency (Confrey & Makar, 2005; Light et al., 2004), the rationale being that these students, with a little work and a minimum investment of resources, can be brought over the proficiency threshold, thereby improving the school's progress metric in the accountability system. School staff fail to appreciate that given measurement error, many of these students would be expected to score above that cut point on a second testing without any intervention, and students who had scores just above the cut point have a similar likelihood of scoring *below* the cut on the next testing if there is no growth in their underlying proficiency (Confrey & Makar, 2005). More important, the focus on the "bubble kids" results in those farthest behind being neglected.

Research conducted thus far has tended to focus on identifying supporting conditions and implementation challenges for data-driven decision making (Marsh, Pane, & Hamilton, 2006). There is a pronounced lack of data on the prevalence of the use of these data systems and the ways in which system use is influencing the instruction received by students. Comparisons among school districts have found that those districts making more progress in terms of improving student achievement are also investing more effort in using data to inform decisions (Snipes, Doolittle, & Herlihy, 2002), but rigorous evidence with respect to the causal relationship, if any, between system-level data-driven decision making and student achievement is lacking.

Technology Supports for Assessments for Accountability

The movement toward standards-based reform that has gained increasing purchase within American education since the mid-1980s has emphasized the importance of articulating standards for the content that students should learn and the level of proficiency they should demonstrate. States and districts not only articulate these standards but also hold students, teachers, schools, and districts accountable for meeting them, a process that necessarily involves student assessment.

The NCLB legislation has given strong impetus to the increased use of assessments tied to standards. The requirement for annual student testing tied to state standards and the demonstration of annual yearly progress for students overall and for student subgroups has created a demand for efficient assessment tools that teachers and administrators can use to gauge student progress and identify potential trouble spots. If teachers regularly and efficiently test students on state standards, the reasoning goes, they will be able to use those test results in their lesson planning to focus instruction for both the whole class and for individual students. Many educators and policymakers see a role for technology in supporting assessments for accountability. By facilitating teacher assessment of student performance on standards, technology-based systems can strengthen each teacher's instructional focus and standardize the content addressed across classrooms, schools, and districts within a state (Means, 2006).

Advantages of Technology-Based Assessment Systems

Technology-based assessment systems offer a number of advantages. Standards-based assessments can be easily customized. Commercial vendors have large banks of test items mapped to specific skills and tests can be developed that fit a specific state's or district's standards for a particular grade level. Some systems allow teachers to modify items or add their own. If students take the assessments online, another advantage is rapid scoring. Rather than having to ship test answer sheets to a vendor for scoring and waiting for months to get reports of individual student performance, the teacher can get individual student and whole-class profiles nearly instantaneously. Administrators at the school and district level can get a similarly rapid and detailed look at how students are performing against standards. Some technology-based assessments are promoted as "benchmark tests," even providing predictions of the likelihood that a student or student group will attain proficiency on the end-of-year state assessment.

This increased emphasis on assessment for accountability purposes has stimulated the market for commercial, computer-based assessment systems. Accountability-related testing systems have been part of what *Education Week* called "the greatest pre-collegiate testing boom in history." By 2004, education technology directors in 16 states reported that their states were offering Internet- or computer-based assessments of student achievement (Bakia, Mitchell, & Yang, 2007). In 2005–06, the number of states offering computer-based assessments had increased to 22 (*Education Week*, 2006). Of these 22 states, 13 make current state assessment results available through a centralized portal, 11 make current state assessment subscale or item results available, 13 provide performance data over time, and 17 link individual identifiers to assessment results. Table 1 identifies a number of computer-based testing systems, briefly describes each system, and notes the system's stated purpose.

In addition to the technology-based assessments being used at the district and state levels, technology is being used increasingly in national and international assessment systems. For example, the science framework for the 2009 National Assessment of Educational Progress and the 2006 and 2009 cycles of the Program for International Student Assessment are piloting computer-based items or entire test forms.

TABLE 1
LIST OF TECHNOLOGY-SUPPORTED ASSESSMENTS FOR ACCOUNTABILITY PURPOSES

Assessment Product	Developer	Description	Purpose
Texas Math and Science Diagnostic System	State of Texas	Web based. Math component for grades 3–12, science grades 4–12. Benchmark assessment given at beginning of semester and at regular intervals thereafter. Content areas: geometry, spatial reasoning, measurement, probability and statistics, patterns, relationships, algebraic thinking (math), elementary and middle school science, high school physics, chemistry, and biology.	Test practice in mathematics and science for use in student remediation and acceleration.
FCAT Explorer	State of Florida	Web based. Test and skills practice keyed to FCAT. Grades 3 through 11. Content areas: math, science, reading, writing. Accessible by educators, parents, and students.	Test and skills practice, performance feedback, learning guidance, skill reinforcement, remediation, enrichment.
Progress Assessment Series	Pearson Education	Web based. Formative assessment series. Grades 3–8. Content areas: reading and math. Lexile scale used for reading, quantile scale for math. Reading component has pretest followed by three progress-monitoring tests, math component has pretest and six progress-monitoring tests designed for completion in one class period.	To measure student progress, tie assessment to classroom instruction, and forecast student performance on state-specific proficiency standards.

Pinnacle Plus	Excelsior Software	Web based. Integrated data system that provides educators, parents, and students with information on academic progress, attendance, grade reporting.	To assess student performance with multiple grading and weighting schemes and link performance to standards and benchmarks. To track and analyze student performance at the individual, class, and school level.
LearnerLink	HOSTS Learning	Web based. Provides resources to assist teachers in resource management, alignment of instruction to standards and benchmarks, diagnosis of student deficiencies or needs, prescriptive lesson planning and instruction, and access to data for decision making. Provides access to an assortment of assessment instruments and enrichment activities.	To provide teachers with formative assessment and lesson-planning assistance to align instruction with state/local standards and large-scale standardized tests. Provides access to instructional resources and assessment instruments. Provides data to identify student needs. Provides tools for resource and time management.
EduTest	PLATO	Web based. Grades K–8. Standards-based formative and benchmark assessment system. Assessment instruments may be customized to suit school and district needs. Data provided to educators to disaggregate and make decisions.	Classroom formative assessment and district benchmark assessments that provide midyear information on student progress with respect to NCLB requirements for annual improvement. Data provided for tracking student progress and needs and for decision making.

FCAT, Florida Comprehensive Assessment Test.
HOSTS, Helping One Student To Succeed.
NCLB, No Child Left Behind.

Empirical Findings

So far, the results of such investments show mixed returns. On the one hand, states have documented efficiencies in scoring, task presentation, and delivery linked to the use of online assessments. At the same time, few of the operational accountability systems, if any, are using the types of technology-based items that would leverage the capabilities of the technology to present complex problem-solving tasks such as animations and simulations. There also remain unanswered questions about whether the format of delivery (technology versus pencil–paper) affects students' academic performance in ways that disadvantage students who have limited access to technology for practice.

There have been laboratory-based studies of the comparability of paper–pencil and technology-based test forms that suggest why it is important to examine the effects of format on test performance. Since the 1990s, researchers have conducted rigorous and systematic studies of the features of assessments and their administration that might affect student performance when assessments transition from one presentation modality to another (Paek, 2005). Typically, researchers have examined the relationship of three variables to students' performances: (1) students' freedom to review and revise responses during the assessment; (2) features of the graphics and text on computer screens; and (3) students' familiarity with computers (Russell, Goldberg, & O'Connor, 2003). Such studies are conducted to reveal whether the psychometric qualities of the items and tests are affected when the mode of delivery changes from paper–pencil to technology-based. This type of research examines the differences in test-score and individual-item means and standard deviations, with little concern about underlying construct validity issues. Nor does this line of research examine comparisons of student achievement in schools or districts before and after implementing technology-based assessment systems.

At first, as technology entered the assessment world, the transition of a large-scale or standardized test from a paper–pencil administration to a technology-supported delivery was marked by few, if any, differences in the nature and types of items and tasks presented. In fact, the technology-supported version of the assessment was often exactly the same as the paper–pencil version except that the items and responses were presented on a computer screen. However, in the past 5–7 years, the science and art of technology-based assessment design and development has evolved to include new forms of assessment tasks such as simulations and animations that present visually rich stimuli, dynamic representations, inventive response formats, and the opportunity to assess different and more complex knowledge and skills (Quellmalz & Haertel, 2005). Technology has freed assessment designers from the constraints of only being able to present items on paper in a static format. With technology, assessment tasks and items can be presented dynamically, which greatly increases the kinds of knowledge and skills that can be assessed (Means & Haertel, 2003).

Opportunities to draw on new technologies to construct scenarios, interact with examinees, capture and evaluate their performances, and model the information they provide has opened a new era of assessment design and development. Computers and other media offer potential solutions to the practical challenges of assessing complex constructs such as model-based reasoning and science inquiry in the context of challenging content. Technology confers advantages in

- measuring phenomena that cannot easily be observed in real time, such as seeing things in slow motion (e.g., a wave) or speeded up (e.g., erosion). It is possible to freeze action or replay it.
- modeling phenomena that are invisible to the naked eye (e.g., movement of molecules in a gas).
- working safely in the simulation of a lab that would otherwise be hazardous (e.g., using dangerous chemicals) or messy and time-consuming in an assessment.
- conducting several repetitions of an experiment in limited assessment time while varying the parameters (e.g., rolling a ball down a slope while varying the mass, angle of inclination, or coefficient of friction of the surface).
- manipulating objects to solve problems or express solutions, such as moving concept terms and relationship labels in a concept map.

Interactive assessment tasks have been developed that demonstrate the advantages of technology-based assessments for assessing higher order knowledge and skills. Bennett and his colleagues at Educational Testing Service (Bennett, Jenkins, Persky, & Weiss, 2003), for example, developed a simulation of the physics of a hot air balloon that employed an interactive response format. Students are able to design and conduct experiments and interpret results. Researchers at SRI (Quellmalz et al., 2005) developed two computer-based simulations that assess science content and inquiry skills in the domains of force and motion and ecosystems. These tasks employ interactive response formats, including log files that can be analyzed by computer for evidence of solution paths. Robert Mislevy and his colleagues on the Principled Assessment Designs in Inquiry Project have adapted a well known paper–pencil assessment task, historically referred to as "Mystery Powders," to a technology-based assessment. This assessment, which was designed to assess scientific reasoning in the area of chemistry (Seibert, Hamel, Haynie, Mislevy, & Bao, 2006), employs a simulation and a selected response format in the context of a computerized adaptive test. While all the assessment tasks described earlier were developed with standardized, large-scale assessments in mind, the technologies that support them are also applicable for use in formative assessments (see the discussion of formative assessment that follows).

Technology-Supported Formative Assessments

Although vendors of the kinds of accountability-oriented assessment systems discussed earlier often describe their systems as "formative" in nature, most of the systems actually provide limited information that is usable for guiding instruction. Accountability-related assessments are designed to cover a broad collection of instructional objectives, with only a few items or even just a single item per objective. The performance reports received by teachers reveal which students are doing better and which more poorly on collections of objectives, but typically lack the kind of detailed information that could actually be used to shape future instruction for those who are not doing well. The teacher thus is put in the position of knowing who is struggling but not knowing the basis of their difficulties.

Furthermore, there is concern about the adequacy of the content covered by standardized tests. Measurement experts have pointed out the mismatch between current standardized tests and the needs of classroom teachers (Pellegrino, Chudowsky, & Glaser, 2001; Popham, 1995). Influential reform documents support curricular activities that foster deep understanding in subject areas. This understanding often is associated not only with fluency in the facts, concepts, and principles used in the domain, but also with the capability to think and reason like an expert. Commercial achievement tests that primarily provide evidence of reading comprehension and fact-based recall are insufficient to measure the learning associated with complex forms of reasoning, metacognitive strategies, and multi-step problem solving.

Lee Shulman (2007) distinguished the characteristics of classroom formative assessments from those of assessments for accountability in a recent opinion piece. He wrote:

> Assessment should not only serve as an external evaluation and public conscience . . . at the very least, it should do no harm to instruction, and at best, it should guide, support, and enrich it. . . . Embedded measures will necessarily be designed with a different "grain size" from those designed exclusively for external, high-stakes assessments. They will be more particular than general; more dedicated to measuring individual student progress than institutional success; repeatedly administered rather than being single, end-of-course events; and highly transparent to students and teachers. They will have quick turn-around times rather than providing highly secure, secretive, and delayed feedback of current high-stakes environments. This is assessment as a regular physical exam rather than a public autopsy.

In an influential review of research studies on formative assessment, Black and Wiliam (1998) defined this practice as "all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (p. 2). Thus, formative assessments are designed specifically to inform future instruction (Black & Wiliam). They argue that the "formativeness" of an assessment is found in the test's intersection with the instruction that occurs in classrooms. State-of-the-art thinking about assessment is built on a premise that insights about students' learning will materialize from assessment items and tasks that test deep conceptual understandings in content domains, including the relationships among the concepts (Pellegrino et al., 2001). As a result, formative assessments must go beyond providing information about whether a student has achieved mastery of particular standards or learning objectives. They must provide information about student understandings in ways that teachers can use to improve students' deep understanding of the content. Formative assessments emphasize depth of information about a narrow range of concepts and skills in contrast to end-of-year state assessments, which prioritize breadth of coverage. Black and Wiliam, in their review of the effects of formative assessment on student learning, reported an average effect size of .40. This is a substantial effect and demonstrates the potency of assessments as a lever of positive change in student learning. They concluded: "We know of no other way of raising standards for which such a strong *prima facie* case can be made on the basis of evidence of such large learning gains" (p. 19).

Advantages of Technology Supports for Formative Assessment

The prior section describes attributes and effects of paper–pencil and hands-on performance assessments that have been designed to gather formative information on students' performance. Such approaches, while fruitful, are limited to assessing content and processes that teachers can conduct in their classrooms or that use static pencil–paper formats. The use of technology-based assessments unleashes new possibilities for the content and processes that can be assessed. Technology can provide a means to assess high-level conceptual understanding, reasoning, and skills in formative, classroom-based assessment (Bennett, 2001; Linn, Lee, Tinker, Husic, & Chiu, 2006; Mislevy et al., 2003; Quellmalz & Haertel, 2005).

Technology can play a critical role in helping test developers address both domain and strategic knowledge through the use of assessment tasks that employ rich task environments, innovative response formats, immediate feedback, and reliable and informative scoring. For the purposes of this chapter, technology-supported formative assessments are those assessments delivered by computer to a student or group of students with the intent of providing information that describes the conceptual understandings and strategies that students possess within and across learning domains. The formative information the test provides must be linked to future instruction and the provision of particular learning opportunities. We envision the formative assessment of the future to be (1) inspired by the constructs that have emerged from the learning sciences; (2) designed so that teachers can readily identify gaps in students' learning; (3) delivered via technology; and (4) intimately associated with instruction.

"*Facets.*" An example of technology-supported formative assessment can be offered from the work of Jim Minstrell and his associates. Minstrell has spent years compiling a set of student conceptions about force and motion based upon both the research literature and the observations of teachers. Some of these ideas, or "facets" in Minstrell's terminology, are considered scientifically correct (or at least correct to the degree one would expect at the stage of introductory physics). Others are partially incorrect, and still others are seriously flawed. The goal of Minstrell's "facets assessments" is to elicit student responses that reveal their underlying thinking. Having developed an inventory of knowledge facets, Minstrell and his colleagues proceeded to develop assessment items that would elicit different responses depending upon which facets a student held (Minstrell, 2001). For example, when asked to reason about the weight of objects totally or partially submerged in a liquid, one set of facets concerns separating the effect of a fluid or other medium from the effect of gravity. A student might think that surrounding forces do not exert any pressure on objects; alternatively, he might think that a fluid medium produces an upward pressure only or that the weight of an object is directly proportional to the medium's pressure on it. Some students may have memorized the mathematical formula for net buoyant pressure and may be able to apply it to some problems in order to obtain a correct answer, but might nonetheless lack the facet for a qualitative conceptual understanding (net upward push is because of differences in pressure gradients).

Minstrell and his colleagues (2001, 2003) have developed a computer-based assessment system to get at students' facets. The student is presented with a problem situation and a set of multiple-choice answers, each of which is associated with a specific facet. One example might be

> A solid cylinder is hung by a long string from a spring scale above a container of water. The reading on the scale shows that the cylinder weighs 1.0 lb. About how much will the scale read if the cylinder which weighs 1.0 lb. is submerged just below the surface of the water?

After choosing an answer to the original question, the student is asked to provide the reasoning behind the original answer. The system compares the facet associated with the student's explanation to that associated with the original answer choice. Over multiple problems, the system diagnoses the student's probable facets and the consistency between student predictions and explanations and presents the teacher with reports of this diagnosis and with an instructional prescription appropriate for the diagnosed facets.

The FACETS web site (see http://www.facetinnovations.com/daisy-public-website/daisy/fihome/6.html) demonstrates the interplay between assessments and instruction. In addition to providing teachers with access to the diagnostic assessments, the site provides sample lessons geared to specific misconceptions. The purpose is to encourage students to apply their beliefs to new situations, examine their own reasoning, and see where their expectations are confirmed and where there are discrepancies between their beliefs and what they observe actually happening. By keying these experiences to students' specific misconceptions, teachers can increase the likelihood of conceptual growth.

Simulations. As noted in the prior section on assessments for accountability purposes, several assessment experts have developed interactive assessment tasks that use simulations to assess complex, multistep reasoning in various science domains (see Bennett et al., 2003;

Quellmalz & Kozma, 2003; Quellmalz et al., 2005; Seibert et al., 2006). Such interactive assessments can be used for both accountability and formative purposes. They can be used formatively if they are well aligned to the content being covered in a classroom, if they identify gaps in a student's understanding and skills, and if they can be used to prescribe participation in instructional activities, the review of content taught, or exposure to additional resources. The feasibility of using such interactive assessment tasks to assess students' higher order reasoning for formative purposes is being documented. Assessment data collected using the SRI science inquiry simulations and Bennett's hot air balloon simulation were tested in classrooms and indicate that computer-based simulation tasks can be successfully used to gather information on higher order, problem-solving science skills. In addition, teachers reported that the assessment data collected as part of the SRI science inquiry tasks was useful in informing their instructional decision making.

E-learning. In a final example of how interactive technology-based tasks are being used for formative purposes, we describe the use of such assessment tasks in a global e-learning program that was developed by the Cisco Networking Academies Program. The articles cited in succeeding discussions provide detailed descriptions of how such interactive tasks can be designed using evidence-centered design and how they are presented and delivered online (Williamson, Bauer, Steinberg, Mislevy, & Behrens, 2004); the flexibility and power of the scoring and evaluation rules that can be used to provide useful information on student performances (DeMark & Behrens, 2004; Levy & Mislevy, 2004); and the types of psychometric analyses that can be conducted and reported within a formative assessment system (Levy & Mislevy).

The Networking Performance Skill System project, which supports the work of the Cisco Networking Academies, had as its goal the building of an online performance-based assessment prototype (Behrens, Mislevy, Bauer, Williamson, & Levy, 2004). This interactive prototype, when fully developed, was to be incorporated into the assessment system that served the Networking Academies. The assessment system included the use of online selected response items for formative as well as summative feedback to students and instructors. The selected response items were used in chapter, mid-term, and final examinations. The performance assessment tasks address four broad constructs: troubleshooting processes, implementation of those processes, design of solutions, and declarative knowledge about troubleshooting. The hope was that the use of online performance assessments that made use of open-ended response formats, rich and authentic presentations of situations that occur in the networking environment, and the use of multi-step problems would provide diagnostic feedback as well as situative information that would go well beyond the kinds of information provided by the selected response items. The online assessments that were developed illustrate the kinds of higher order, problem-solving tasks that can be presented using technology (Williamson et al., 2004), the use of natural language processing for understanding complex responses to free-response tasks (DeMark & Behrens, 2004), and the sophisticated measurement models and approaches to evidence accumulation (e.g., trace of interactions) that can be implemented for use with such tasks (Levy & Mislevy, 2004). The current formative system, as implemented by Cisco, does not use all of these features (e.g., natural language processing) but applies some of them. Evidence on the effectiveness of the formative interactive tasks as a means of improving student performances is not yet available, but there are plans to collect such data and share results.

Classroom Communication Systems

The vision of formative assessment painted earlier is one where assessment and instruction are intricately interwoven. Formative assessments probe students' understanding of the concepts being taught and provide the basis for designing further instruction for those concepts. This vision is consonant with the recommendations of learning scientists (Bransford, Brown, & Cocking, 2000; Pellegrino et al., 2001) and would be a major advancement in classroom instruction if consistently enacted. However, some learning technology researchers want to go a step farther—to harness the power of technology to greatly compress the timeframe of instruction–assessment–instruction cycles into minutes rather than days. They propose to do this by "instrumenting" the classroom with a communication system that can embed mini, just-in-time assessments into instruction (Crawford, Schlager, Penuel, & Toyama, in press).

Classroom communications systems consist of networked sets of computers, personal data assistants, or small wireless input devices. A teacher uses a laptop computer to project a question on a screen, along with answer choices for the question. Every student in the class can answer the question simultaneously, and either through a radio-frequency signal or a wireless network, student answers are aggregated and presented, usually in the form of a histogram. Typically, the correct answers are not displayed, so teachers may reteach, encourage peer discussion about the concept being taught, or ask for explanations before displaying and explaining the correct answer. Student responses are collected efficiently, facilitating a rapid cycle of question-and-answer and immediate feedback (Roschelle, Penuel, & Abrahamson, 2004). More advanced classroom network technologies also allow student input of text and graphical responses to teacher questions. The student input devices for such advanced applications could be either student laptops or graphing calculators.

Empirical Findings

Some of the best empirical evidence of the effectiveness of this technology on improving learning comes from the university-level physics instruction conducted by Eric Mazur and his colleagues (see Crouch & Mazur, 2001; Mazur, 1997). To support the development of conceptual understanding of physics content in his college students, Mazur uses a combination of a student response system and a technique he calls "peer instruction." After lecturing for a short time and posing a conceptual question, Mazur poses a question to his class and has each student use the classroom communication system to register a response. He then has his students work with one or two other students to discuss their answers and provide explanations to each other. After engaging in the peer activity, the same conceptual question is delivered to the class again using the communication system. Typically, the number of correct answers increases dramatically and Mazur proceeds to the next topic. However, if student responses indicate a widespread misconception, Mazur can immediately adjust his instruction to address it. Mazur and his colleagues report a positive association between students' understanding of introductory physics concepts and the use of a student response system supplemented with peer instruction (Fagan, Crouch, & Mazur, 2002; Mazur). Students in classrooms using interactive technologies and engagement methods similar to those used in peer instruction also make higher gains on measures of understanding than students in comparison classrooms (Hake, 1998; Sokoloff & Thornton, 1997). But the study designs do not support untangling the contributions of the communications technology as opposed to the nature of the questions posed and the peer instruction (Roselli & Brophy, 2002).

There exists a new class of technologies that rely on classroom networks and that extend the capabilities of classroom communication systems (Penuel & Riel, in press). These new technologies enable students to input a variety of responses, including open text and graphical images. Classroom network technologies are not content independent, as is the case when using student response systems; rather, the technology helps structure student interaction with content in ways that help make difficult concepts visible and more concrete. For example, in one application in mathematics (Hegedus & Kaput, 2004), students are given handheld computers connected to a classroom network, and a teacher's computer assigns them randomly to groups of three students each. The teacher can then assign a problem the students must collaborate to solve, such as matching a fraction to a decimal value. Each student has one fraction and decimal, and then uses her or his handheld device to "give" one of the two to a peer to create a match.

As a result of simulations and other forms of classroom activities that can be implemented using these devices, distinctive student roles emerge that teachers must be prepared to support. In particular, students become active agents in classroom activities; they are not manipulating the system or process from the outside but from within it (Colella, 2000). For example, in a simulation designed to teach students about how an infectious disease can spread within a population, students may act as individual organisms in the population, using their computers to interact with other organisms (other students). They unwittingly "infect" others with the disease, and like real agents know only whether they themselves are infected and have to infer underlying causes from the pattern of interaction they can recall or see. Specific subject matter plays a critical role in enabling and constraining the kinds of classroom activities that can occur when using network technologies such as these.

To date, there is little research on the effects of classroom networks on student achievement. However, there is some preliminary evidence that participatory simulations that rely on network technology can be effective in promoting learning (Hegedus, 2003; Lonsdale, Baber, & Sharples, 2004; Wilensky & Stroup, 2000). For example, Wilensky and Stroup present evidence that when students construct parameters that enable them to test different conjectures and hypotheses, there is an opportunity for them to come to a deeper conceptual understanding. In a simulation in which students are asked to model traffic flows in an urban area, students can test different hypotheses about the pattern of traffic lights that will produce the least traffic. They may arrive, through their experimentation, at a model that even approximates "timed" lights that have been developed in urban areas to increase flow on certain highly trafficked arteries. One advantage of using classroom communication systems for real-time diagnosis of student understanding is that a record of the performance of each student as well as the class as a whole can be maintained. Thus, detailed data on student learning are available for later analysis and reflection.

The Vision of an Integrated System

We have reviewed four categories of technology use to support assessment and student learning. Two of these categories are supports for different types of assessments—those we termed "accountability related" and "formative" assessments. We have noted that accountability-related assessments are a central tool of standards-based educational improvement and tend to cover broad areas of content and proficiency. While technology can provide advantages in terms of customizing and rapid scoring of large-scale assessments, the nature of these tests themselves limits their utility in guiding instruction. Formative assessments, on the other hand, have the potential to inform teacher choices of what and how to teach. Technology-supported formative assessments provide much more detailed information about student understanding and can do so at a point in time when further instruction can benefit from this information.

The other two categories of technology use discussed in this chapter involve systems for storing, aggregating, and displaying assessment data. In both cases, the purpose is to enable educators to use assessment results as part of their decision making. In the case of student data management decision systems, the data typically are scores on accountability-related assessments and the decision makers may be at the state, district, or school level. The express purpose for using the system is to distribute resources and refine curriculum and instruction in ways that increase student scores on accountability assessments. The assessment data in the student data system are likely to be "refreshed" only once or twice a year, and hence the data are more useful for "macro" education decisions and for planning major curriculum and professional development efforts rather than for day-to-day instructional decision making. The latter type of decision making is the target for classroom communication systems. These systems are designed for classroom-level decision making and are intended to support moment-by-moment adjustments in instructional approach and support.

But these differences in how the systems are typically used today in practice are not inevitable. In an ideal world, the instructionally diagnostic formative assessments and classroom communication system practices described here could be integrated with accountability systems to provide more complete data on students. As noted earlier, most of the student data management systems can incorporate multiple measures of student learning, and there is no *a priori* reason why they could not include results of multiple classroom diagnostic assessments. The stumbling block has been the labor required to put assessment results into this kind of system. The automated collection of formative assessment data that is supported by classroom communication systems has the potential to address this need for data input. Student responses made as part of classroom activities orchestrated through a classroom communication system could be uploaded to a student data management system to provide a detailed record of students' performances over time. Such an integrated system would enable administrators, teachers, parents, and students to review not just scores on accountability assessments but a detailed set of performances in specific content areas. The vignette that follows provides a glimpse of the potential such an integrated system could provide.

Integrating different types of systems is no small task from a technical standpoint, and creating the social and organizational climate that would support this kind of record of instruction and performance is perhaps an even greater challenge. But the commitment to improving all students' learning is exerting pressure in this direction, and technology will be an essential component of any integrated system.

The Integrated System in Practice: A Vignette

Our fictional eighth-grade science teacher begins her day by meeting with the middle school principal. In this meeting, the teacher discusses the performance of last year's students on the statewide science assessment, a discussion that is informed by data from the district's data management system. She then continues her day with several class periods during which she presents new science content and guides her instructional decision making using a classroom communication system. Later in the day she engages a small group of students in a simulation-based assessment on electric circuitry. Her day ends with a technology-based presentation to the parents of her students at the annual "Parents Night" activity. In this presentation, she uses the simulation-based assessment to illustrate the depth of knowledge that children will have to master to meet the standards set by the state. During the presentation, she relies on the district's data management system to provide charts and tables that compare their children with their counterparts in other schools within the district and the state.

Using the District's Data Management System to View Student Performance

Before the beginning of the school day, our teacher meets with the school principal to discuss the performance of last year's eighth-grade students on the statewide science assessment. These students did not perform at what the district considers a satisfactory level on the science exam and, in particular, they were below the district and state averages on items measuring students' understanding of forces and motion as well as on items associated with the application of science inquiry skills. Students had difficulty with items that required the understanding that a variety of forces govern the structure and motion of objects in the universe. To make her points, the middle school principal goes online and accesses the district's data management system, which dynamically produces graphs and charts that compare the total score on the science exam of eighth graders in this school to their counterparts in other schools within the district and the state, and to eighth graders in other schools with similarly large proportions of low-income students.

The principal drills down into the assessment results to compare student performance on particular standards and items. She and the teacher find that while the performance of their students on items that measure knowledge of force and motion was only slightly lower than the district and state averages, there was a significant gap reflected in the performance of this school's eighth graders on the more complex, open-ended science inquiry items. Student performance was slightly below the district average on the items about electric currents and magnets exerting a force on certain objects and each other. However, in science inquiry, most students performed well below the district average on the items that required that they express a testable hypothesis, provide explanations, and make predictions. Using the data management system, the teacher and principal look at the specific items used to measure these science inquiry skills in last year's state test and can see examples of the responses of students from their school as well as some responses that the state has posted to illustrate proficiency on these open-ended items. The teacher realizes that students last year had difficulty distinguishing hypotheses from data and assures the principal that she will give more emphasis to that skill this year. After printing out copies of the assessment items and illustrative "Below Basic" and "Proficient" responses, the teacher starts thinking about her lesson plan for the day and how she can incorporate the insight she has gained in her discussion with the principal into her instruction.

Guiding Instruction Using a Classroom Communication System

In our teacher's first-period class, she presents basic information on force and motion, including definitions and examples of concepts such as acceleration, rate, and velocity. As part of the lesson, she demonstrates each concept using a physical apparatus (e.g., ball and ramp). After the demonstration, the students form small groups and each group conducts two simple experiments with the same materials the teacher used. The teacher uses the technology-based classroom communication system to project questions to which each group must respond. The questions call for predictions about what will happen in each experimental trial. The teacher takes care to use the term "hypothesis," which seemed to confuse some of last year's students when they took the state test. Each group transmits its answer to her question (i.e., their prediction) via the communication system, which posts the answers for the entire class to view. The groups then discuss the rationales they had for their conflicting hypotheses before following the protocol for using the balls and ramp, recording the results on their handheld computers and transmitting them to the teacher's laptop.

The teacher engages the class in a discussion about how their data compare to their hypotheses and why some of them were surprised by what happened. She notes with relief that these students seem to have a pretty good understanding of the difference between hypotheses and data and how examining data can lead to new hypotheses. But the predictions her students made and her discussion of their observations and what they would do next suggest that many of them are missing some key understandings in the physics of motion. Surprisingly, many of these students do not recognize that unbalanced forces acting on an object change both the object's speed and direction. She accesses an online item bank to pull up questions on this relationship, knowing that the questions in the online bank have been classroom-tested and are designed to reveal different physics misconceptions that research has shown students often have. She poses these questions to her students and has them respond individually using their response systems. A first quick look at the aggregated student responses tells her that only 40% of the class has mastered this key concept about motion. Looking at the student-by-question matrix of responses, the teacher can also see several clusters of students sharing the same misunderstanding. She puts an appropriate flag next to the online learning plan for each of these students and concludes that additional class time will need to be spent on the physics of motion, making a mental note to look up learning activities recommended for the misconceptions that she discovered at home tonight.

Assessing Inquiry Skills With Simulation-Based Assessment

In the late afternoon, the teacher meets with a small group of students who have been working on an independent science project focused on electrical circuitry. She believes that these students will be motivated by using a piece of software that simulates electrical circuits and responds dynamically when the student introduces a change in current, circuitry, or components (e.g., adds a resistor). Each student gets on a computer to work with the simulation. The students use the simulation to construct a model that accounts for how electric current moves in open and closed circuits in response to changing parameters. The software prompts them to predict the results of each change they choose to introduce, to analyze what happens when changes are introduced, and to verify quantitatively the scientific laws illustrated. The students explore circuit behavior enthusiastically, frequently calling on their friends to see what they have done and arguing good-naturedly about what will happen when the next change is introduced.

Students do not think of the activity as an assessment, but in fact the system is capturing a detailed trace of each student's work. Based on the assessment, both the teacher and the student can see how well the student understands the role of each parameter that influences electrical circuits and different ways of optimizing their values. The teacher moves around the room, looking over each student's shoulder and using what she has learned from the system's teacher feedback view to help frame questions tailored to the present knowledge state of the individual student, suggesting things they can try with the simulation.

Communicating With Parents Using Technology-Based Tools

At the end of her long day, our fictional teacher has one more professional responsibility—making a presentation to the parents of her students that describes the physical science curriculum in which their children will engage and the status of their children's knowledge of physical science. Our teacher begins by presenting a scope and sequence chart of the physical science curriculum that she is expected to teach this year. In addition to the range of topics that will be covered and the sequence in which they will be introduced, the teacher indicates the percent of time that she will devote to each topic. Having presented this overview, she wants to impress parents about the depth of knowledge and skills in physical sciences that students in her classes will be expected to attain. To make her point, she displays the simulation-based assessment on electrical circuitry that she used with the small group of students earlier in the day and poses some of the assessment questions to the parents. Her students' parents laugh nervously as they realize that they do not understand some of the concepts their eighth graders will be assessed on during state science testing.

After showing parents the simulation-based assessment items, the teacher accesses the district's data management system. She presents several histograms that show the performance of her eighth-grade classes from last year on the statewide exam. In response to a parent question, she shows how their performance compared with eighth graders from other schools in the district and with eighth graders from schools with similar populations. She illustrates in detail the science inquiry skill areas where last year's students did not perform as well as would be expected. She then switches back to the online assessment system to show where this year's eighth graders were with these skills at the first of the year as compared with today, when they engaged in the force and motion lesson. She is proud to be able to show that this year's students are already ahead of where last year's eighth graders were at spring testing.

References

Bakia, M., Mitchell, K., & Yang, E. (2007). *State strategies and practices for educational technology*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development.

Behrens, J.T., Mislevy, R.J., Bauer, M., Williamson, D.M., & Levy, R. (2004). Introduction to evidence centered design in a global e-learning program. *International Journal of Testing*, *4*(4), 295–301.

Bennett, R.E. (2001, February). How the Internet will help large-scale assessment reinvent itself. *Education Policy Analysis Archives*, *9*(5). Retrieved July 12, 2007, from http://epaa.asu.edu/epaa/v9n5.html

Bennett, R.E., Jenkins, F., Persky, H., & Weiss, A. (2003). Assessing complex problem-solving performances. *Assessment in Education*, *10*, 347–359.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment and Education*, *5*(1), 7–74.

Bransford, J.D., Brown, A.L., & Cocking, R.B. (2000). *How people learn: Brain, mind, and experience*. Washington, DC: National Academies Press.

Colella, V. (2000). Participatory simulations: Building collaborative understanding through immersive dynamic modeling. *The Journal of the Learning Sciences*, *9*(4), 471–500.

Confrey, J., & Makar, K.M. (2005). Critiquing and improving the use of data from high-stakes tests with the aid of dynamic statistics software. In C. Dede, J.P. Honan, & L.C. Peters (Eds.), *Scaling up success: Lessons learned from technology-based, educational improvement* (pp. 198–226). San Francisco: Jossey-Bass.

Crawford, V.M., Schlager, M., Penuel, W.R., & Toyama, Y. (in press). Supporting the art of teaching in a data-rich, high performance learning environment. In E.B. Mandinach & M. Honey (Eds.), *Linking data and learning*. New York: Teachers College Press.

Cromey, A. (2000, November). *Using student assessment data: What can we learn from schools?* (Policy Issues No. 6). Oak Brook, IL: North Central Regional Educational Laboratory.

Crouch, C.H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *The Physics Teacher*, *69*, 970–977.

DeMark, S., & Behrens, J.T. (2004). Using statistical natural language processing for understanding complex responses to free-response tasks. *International Journal of Testing*, *4*(4), 371–390.

Deming, W.E. (1986). *Out of crisis*. Cambridge, MA: MIT Center for Advanced Engineering Study.

Education Week (2006). *Technology Counts 2006: A special state-focused supplement to Education Week*. Bethesda, MD: Editorial Projects in Education.

Fagan, A.P., Crouch, C.H., & Mazur, E. (2002). Peer instruction: Results from a range of classrooms. *The Physics Teacher*, *40*(4), 206–207.

Feldman, J., & Tung, R. (2001). Using data based inquiry and decision-making to improve instruction. *ERS Spectrum*, *19*(3), 10–19.

Hake, R.R. (1998). Interactive-engagement versus traditional methods. *American Journal of Physics*, *66*, 64–74.

Hegedus, S. (2003, July). *Improving algebraic thinking through a connected SimCalc Mathworlds classroom*. Paper presented at the 27th Conference of the International Group for the Psychology of Mathematics Education held jointly with the 25th Conference of the PME-NA, Honolulu, HI.

Hegedus, S., & Kaput, J. (2004, September). *An introduction to the profound potential of connected algebra activities: Issues of representation, engagement and pedagogy*. Paper presented at the 28th Conference of the International Group for the Psychology of Mathematics Education, Bergen, Norway.

Herman, J., & Gribbons, B. (2001). *Lessons learned in using data to support school inquiry and continuous improvement: Final report to the Stuart Foundation*. Los Angeles: UCLA Center for the Study of Education.

Levy, R., & Mislevy, R.J. (2004). Specifying and refining a measurement model for a computer-based interaction assessment. *International Journal of Testing*, *4*(4), 333–369.

Light, D., Wexler, D., & Heinze, J. (2004, April). *How practitioners interpret and link data to instruction: Research findings on New York City schools' implementation of the Grow Network*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Linn, M.C., Lee, H.S., Tinker, R., Husic, F., & Chiu, J.L. (2006). Supporting online material for teaching and assessing knowledge integration in science. *Science*, *313*, 1049–1050.

Lonsdale, P, Baber, C., & Sharples, M. (2004, September). *Engaging learners with everyday technology: A participatory simulation using mobile phones*. Paper presented at the Mobile Human Computer Interaction 2004: 6th International Symposium, Glasgow, UK.

Marsh, J.A., Pane, J.F., & Hamilton, L.S. (2006). *Making sense of data-driven decision making in education: Evidence from recent RAND research*. Santa Monica, CA: RAND Corporation.

Mason, S. (2002, April). *Turning data into knowledge: Lessons from six Milwaukee public schools*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ: Prentice Hall.

Means, B. (2006). Prospects for transforming schools with technology-supported assessment. In R.K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 505–521). New York: Cambridge University Press.

Means, B., & Haertel, G.D. (Eds.). (2003). *Evaluating educational technology: Effective research designs for improving learning*. New York: Teachers College Press.

Minstrell, J. (2001). Facets of students' thinking: Designing to cross the gap from research to standards-based practice. In K. Crowley, C.D. Schunn, & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 415–443). Mahwah, NJ: Erlbaum.

Minstrell, J. (2003). *Facets of learning*. Seattle, WA: FACETS Innovations.

Mislevy, R., Hamel, L., Fried, R.G., Gaffney, T., Haertel, G., Hafter, A. et al. (2003). *Design patterns for assessing science inquiry* (PADI Technical Report 1). Menlo Park, CA: SRI International.

National Center for Education Statistics (NCES). (2005). *Internet access in U.S. public schools and classrooms: 1994–2003* (NCES 2005-015). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Paek, P. (2005). *Recent trends in comparability studies* (PEM Research Report 05-05). Iowa City, IA: Pearson Educational Measurement.

Pellegrino, J.W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Penuel, W.R., & Riel, M. (in press). The new science of networks and the challenge of school change. *Phi Delta Kappan*.

Popham, W.J. (1995). *Classroom assessment: What teachers need to know*. Boston: Allyn and Bacon.

Quellmalz, E.S., & Haertel, G.D. (2005). *Use of technology-supported tools for large scale science assessment: Implications for assessment practice and policy at the state level*. Washington, DC: National Research Council.

Quellmalz, E.S., & Kozma, R. (2003). Designing assessments of learning with technology. *Assessment in Education*, *10*(3), 389–407.

Quellmalz, E.S., DeBarger, A.H., Haertel, G.D., Schank, P., Buckley, B., Gobert, J. et al. (2005, November). *Exploring the role of technology-based simulations in science assessment: The Calipers Project*. Paper presented at the annual meeting. of the National Science Teachers Association, Chicago, IL.

Roschelle, J., Penuel, W.R., & Abrahamson, A.L. (2004). The networked classroom. *Educational Leadership*, *61*(5), 50–54.

Roselli, R.J., & Brophy, S. (2002, June). *Exploring an electronic polling system for the assessment of student progress in two biomedical engineering courses*. Paper presented at the annual conference and exposition of the American Society for Engineering Education, Montreal, Quebec.

Russell, M., Goldberg, A., & O'Connor, K. (2003). Computer-based testing and validity: A look back into the future. *Assessment in Education: Principles Policy & Practice*, *10*(3), 279–293.

Salpeter, J. (2004, March). Data: Mining with a mission. *Technology & Learning*, *24*(8), p. 30. Retrieved August 27, 2007 from http://www.techlearning.com/showArticle.php?articleID=18311595.

Schmoker, M., & Wilson, R.B. (1995). Results: The key to renewal. *Educational Leadership*, *51*(1), 64–65.

Seibert, G., Hamel, L., Haynie, K., Mislevy, R., & Bao, H. (2006). *Mystery powders: An application of the PADI design system using the four-process delivery system* (Draft PADI Technical Report 15). Menlo Park, CA: SRI International.

Shulman, L.S. (2007, January/February). Counting and recounting: Assessment and the quest for accountability. *Change*. Retrieved April 27, 2007, from http://www.carnegiefoundation.org/change/sub.asp?key=97&subkey=2169

Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement*. Washington, DC: MDRC and the Council of Great City Schools.

Sokoloff, D.R., & Thornton, R.K. (1997). Using interactive lecture demonstrations to create an active learning environment. In E.F. Redish & J.S. Rigden (Eds.), *The changing role of physics departments in modern universities: Proceedings of ICUPE* (pp. 1061–1074). College Park, MD: The American Institute of Physics.

Thorn, C.A. (2002, April). *Data use in the classroom: The challenges of implementing data-based decision-making at the school level*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

U.S. Department of Education (2002, September 24). *Internet access in U.S. public schools up for seventh straight year*. Press release. Retrieved May 27, 2007, from http://www.ed.gov/news/pressreleases/2002/09/09242002b.html

U.S. Department of Education, Office of Educational Technology (2004). *Toward a new golden age in American education: How the Internet, the law and today's students are revolutionizing expectations*. Washington, DC: Author.

Wayman, J.C. (2005). Involving teachers in data-based decision-making: Using computer data systems to support teacher inquiry and reflection. *Journal of Education for Students Placed at Risk*, *10*(3), 295–308.

Wayman, J.C., & Stringfield, S. (2006). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education*, *112*(4), 549–571.

Wayman, J.C., Stringfield, S., & Yakimowski, M. (2004). *Software enabling school improvement through analysis of student data* (Report No. 67). Baltimore: Center for Research on the Education of Students Placed at Risk, Johns Hopkins University.

Wilensky, U., & Stroup, W.M. (2000, June). *Networked gridlock: Students enacting complex dynamic phenomena with the HubNet architecture*. Paper presented at the fourth annual International Conference of the Learning Sciences, Ann Arbor. MI.

Williamson, O.M., Bauer, M., Steinberg, L.S., Mislevy, R.J., & Behrens, J.T. (2004). Design rationale for a complex performance assessment. *International Journal of Testing*, *4*(4), 303–332.