Three themes underlie this chapter’s exploration of how information technologies can advance the design of assessments, based on a merging of the cognitive and measurement advances reviewed in Part II.

- Technology is providing new tools that can help make components of assessment design and implementation more efficient, timely, and sophisticated. We focus on advances that are helping designers forge stronger connections among the three elements of the assessment triangle set forth in Chapter 2. For instance, technology offers opportunities to strengthen the cognition-observation linkage by enabling the design of situations that assess a broader range of cognitive processes than was previously possible, including knowledge-organization and problem-solving processes that are difficult to assess using traditional, paper-and-pencil assessment methods.

- Technology offers opportunities to strengthen the cognitive coherence among assessment, curriculum, and instruction. Some programs have been developed to infuse ongoing formative assessment into portions of the current mathematics and science curriculum. Other projects illustrate how technology fundamentally changes what is taught and how it is taught. Exciting new technology-based learning environments now being designed provide complete integration of curriculum, instruction, and assessment aimed at the development of new and complex skills and knowledge.

- The chapter concludes with a possible future scenario in which cognitive research, advances in measurement, and technology combine to spur a radical shift in the kinds of assessments used to assist learning, measure student attainment, evaluate programs, and promote accountability.
Information Technologies: Opportunities for Advancing Educational Assessment

Technology has long been a major force in assessment. The science of measurement took shape at the same time that technologies for standardization were transforming industry. In the decades since, test designers and measurement experts have often been among the early advocates and users of new technologies. The most common kinds of assessments in use today are, in many ways, the products of technologies that were once cutting-edge, such as automated scoring and item-bank management.

Today, sophisticated information technologies, including an expanding array of computing and telecommunications devices, are making it possible to assess what students are learning at very fine levels of detail, from distant locations, with vivid simulations of real-world situations, and in ways that are barely distinguishable from learning activities. However, the most provocative applications of new technologies to assessment are not necessarily those with the greatest sophistication, speed, or glitz. The greater potential lies in the role technology could play in realizing the central ideas of this report: that assessments should be based on modern knowledge of cognition and its measurement, should be integrated with curriculum and instruction, and should inform as well as improve student achievement. Currently, the promise of these new kinds of assessments remains largely unfulfilled, but technology could substantially change this situation.

Within the next decade, extremely powerful information technologies will become ubiquitous in educational settings. They are almost certain to provoke fundamental changes in learning environments at all levels. Indeed, some of these changes are already occurring, enabling people to
conjecture about their consequences for children, teachers, policy makers, and the public. Other applications of technology are beyond people’s speculative capacity. A decade ago, for example, few could have predicted the sweeping effects of the Internet on education and other segments of society. The range of computational devices and their applications is expanding at a geometric rate, fundamentally changing how people think about communication, connectivity, and the role of technology in society (National Research Council [NRC], 1999b).

The committee believes new information technologies can advance the design of assessments based on a merger of the cognitive and measurement sciences. Evidence in support of this position comes from several existing projects that have created technology-enhanced learning environments incorporating assessment. These prototype cases also suggest some future directions and implications for the coupling of cognition, measurement, and technology.

Two important points of clarification are needed about our discussion of the connections between technology and assessment. First, various technologies have been applied to bring greater efficiency, timeliness, and sophistication to multiple aspects of assessment design and implementation. Examples include technologies that generate items; immediately adapt items on the basis of the examinee’s performance; analyze, score, and report assessment data; allow learners to be assessed at different times and in distant locations; enliven assessment tasks with multimedia; and add interactivity to assessment tasks. In many cases, these technology tools have been used to implement conventional theories and methods of assessment, albeit more effectively and efficiently. Although these applications can be quite valuable for various user groups, they are not central to this committee’s work and are therefore not discussed here. Instead, we focus on those instances in which a technology-based innovation or design enhances (1) the connections among the three elements of the assessment triangle and/or (2) the integration of assessment with curriculum and instruction.

The second point is that many of the applications of technology to learning and assessment described in this chapter are in the early stages of development. Thus, evidence is often limited regarding certain technical features (e.g., reliability and validity) and the actual impact on learning. The committee believes technological advances such as those described here have enormous potential for advancing the science, design, and use of educational assessment, but further study will clearly be needed to determine the effectiveness of particular programs and approaches.
NEW TOOLS FOR ASSESSMENT DESIGN AND IMPLEMENTATION

Computer and telecommunications technologies provide powerful new tools for meeting many of the challenges inherent in designing and implementing assessments that go beyond conventional practices and tap a broader repertoire of cognitive skills and knowledge. Indeed, many of the design principles and practices described in the preceding chapters would be difficult to implement without technology. For purposes of discussing these matters, a useful frame of reference is the assessment triangle introduced in Chapter 2.

The role of any given technology advance or tool can often be differentiated by its primary locus of effect within the assessment triangle. For the link between cognition and observation, technology makes it possible to design tasks with more principled connections to cognitive theories of task demands and solution processes. Technology also makes it possible to design and present tasks that tap complex forms of knowledge and reasoning. These aspects of cognition would be difficult if not impossible to engage and assess through traditional methods. With regard to the link between observation and interpretation, technology makes it possible to score and interpret multiple aspects of student performance on a wide range of tasks carefully chosen for their cognitive features, and to compare the resulting performance data against profiles that have interpretive value. In the sections that follow we explore these various connections by considering specific cases in which progress has been made.

Enhancing the Cognition-Observation Linkage

Theory-Based Item Generation

As noted in Chapter 5, a key design step is the generation of items and tasks that are consistent with a model of student knowledge and skill. Currently, this is usually a less-than-scientific process because many testing programs require large numbers of items for multiple test forms. Whether the items are similar in their cognitive demands is often uncertain. Computer programs that can automatically generate assessment items offer some intriguing possibilities for circumventing this problem and improving the linkage between cognitive theory and observation. The programs are based on a set of item specifications derived from models of the knowledge structures and processes associated with specific characteristics of an item form. For example, the Mathematics Test Creation Assistant has been programmed...
with specifications for different classes of mathematics problem types for which theory and data exist on the processes involved and the sources of solution difficulty (see Figure 7-1 for an example). The computer program automatically generates multiple variations of draft assessment items that are similar to each other in terms of problem types, but sometimes different in terms of the semantic context that frames the problem. Test designers then review, revise, and select from the draft items. Such methodologies hold promise for promoting test design that is more systematic and cognitively principled (Bennett, 1999).

There are, however, challenges associated with ensuring thoughtful generation and use of items employing such automated methods. Sometimes simply changing the semantic context of an item (e.g., from a business scenario to a baseball scenario) can fundamentally change the knowledge that is activated and the nature of the performance assessed. A shift in the prob-
lem context can create knowledge and comprehension difficulties not intended as a source of solution variance. The need for sensitivity to semantic context is but one illustration of the importance of recursively executing the design stages described in Chapter 5, especially the process of validating the intended inferences for an item form across the multiple instances generated.

**Concept Organization**

As noted in Chapter 3, one of the most important differences between experts and novices lies in how their knowledge is organized. Various attempts have been made to design assessment situations for observing learners' knowledge representations and organization (see e.g., Mintzes, Wandersee, and Novak, 2000). An example can be found in software that enables students, working alone or in groups, to create concept maps on the computer. Concept maps are a form of graphical representation in which students arrange and label nodes and links to show relationships among multiple concepts in a domain; they are intended to elicit students' understanding of a domain's conceptual structure (Edmondson, 2000; Mintzes et al., 2000; O'Neil and Klein, 1997; Shavelson and Ruiz-Primo, 2000).

Software developed by O'Neil and Klein provides immediate scoring of students' concept maps based on the characteristics found in concept maps developed by experts and provides feedback to the students. The researchers are also developing an approach that uses networked computers to capture how effectively students work in a team to produce such maps by analyzing such characteristics as the students' adaptability, communication skills, timely organization of activities, leadership, and decision-making skills. The researchers have informally reported finding correlations of around $r = 0.7$ between performance on individually produced concept maps and an essay task (NRC, 1999a).

This concept map research offers just one example of how technology can enhance the assessment of collaborative skills by facilitating the execution of group projects and recording individual involvement in group activities. The ThinkerTools Inquiry Project (White and Frederiksen, 2000) described in Chapter 6 and later in this chapter provides another such example.

**Complex Problem Solving**

Many aspects of cognition and expertise have always been difficult to assess given the constraints of traditional testing methods and formats. For example, it is difficult to assess problem-solving strategies with paper-and-pencil formats. Although conventional test formats can do many things well, they can present only limited types of tasks and materials and provide little
or no opportunity for test takers to manipulate information or record the sequence of moves used in deriving an answer. This limitation in turn restricts the range of performances that can be observed, as well as the types of cognitive processes and knowledge structures about which inferences can be drawn.

Technology is making it possible to assess a much wider range of important cognitive competencies than was previously possible. Computer-enhanced assessments can aid in the assessment of problem-solving skills by presenting complex, realistic, open-ended problems and simultaneously collecting evidence about how people go about solving them. Technology also permits users to analyze the sequences of actions learners take as they work through problems and to match these actions against models of knowledge and performance associated with different levels of expertise.

One example of this use of technology is for the assessment of spatial and design competencies central to the discipline of architecture. To assess these kinds of skills, as well as problem-solving approaches, Katz and colleagues (Katz, Martinez, Sheehan, and Tatsuoka, 1993) developed computerized assessment tasks that require architecture candidates to use a set of tools for arranging or manipulating parts of a diagram. For instance, examinees might be required to lay out the plan for a city block. On the bottom of the screen are icons representing various elements (e.g., library, parking lot, and playground). Explicit constraints are stated in the task. Examinees are also expected to apply implicit constraints, or prior knowledge, that architects are expected to have (e.g., a playground should not be adjacent to a parking lot). The computer program collects data as examinees construct their solutions and also records the time spent on each step, thus providing valuable evidence of the examinee’s solution process, not just the product.

To develop these architecture tasks, the researchers conducted studies of how expert and novice architects approached various design tasks and compared their solution strategies. From this research, they found that experts and novices produced similar solutions, but their processes differed in important ways. Compared with novices, experts tended to have a consistent focus on a few key constraints and engaged in more planning and evaluation with respect to those constraints. The task and performance analyses led the researchers to design the tasks and data collection to provide evidence of those types of differences in solution processes (Katz et al., 1993).

Several technology-enhanced assessments rely on sophisticated modeling and simulation environments to capture complex problem-solving and reasoning skills. An example is the Dental Interactive Simulation Corporation (DISC) assessment for licensing dental hygienists (Mislevy, Steinberg, Breyer, Almond, and Johnson, 1999). A key issue with assessments based on simulations is whether they capture the critical skills required for successful
performance in a domain. The DISC developers addressed this issue by basing the simulations on extensive research into the ways hygienists at various levels of expertise approach problems.

The DISC computerized assessment is being developed for a consortium of dental organizations for the purpose of simulating the work performed by dental hygienists as a means of providing direct evidence about how candidates for licensure would interact with patients. The foundation for the assessment was a detailed analysis of the knowledge hygienists apply when they assess patients, plan treatments, and monitor progress; the analysis was derived from interviews with and observations of several expert and competent hygienists and novice students of dental hygiene. Thus, the initial phase of the effort involved building the student model by using some of the methods for cognitive analysis described in Chapter 3.

The interactive, computer-based simulation presents the examinee with a case study of a virtual patient with a problem such as bruxism (chronic teeth grinding). The simulation provides evidence about such key points as whether the examinee detects the condition, explores connections with the patient’s history and lifestyle, and discusses implications. Some information, such as the patient’s medical history questionnaire, is provided up front. Other information, such as radiographs, is made available only if the examinee requests it. Additional information stored in the system is used to perform dynamic computations of the patient’s status, depending on the actions taken by the examinee.

This mode of assessment has several advantages. It can tap skills that could not be measured by traditional licensing exams. The scenarios are open-ended to capture how the examinee would act in a typical professional situation. And the protocols are designed to discern behaviors at various levels of competency, based on actual practices of hygienists.

MashpeeQuest is an example of an assessment designed to tap complex problem solving in the K-12 education context. As described by Mislevy, Steinberg, Almond, Haertel, and Penuel (2000), researchers at SRI International have developed an on-line performance task to use as an evaluation tool for Classroom Connect’s AmericaQuest instructional program. One of the goals of AmericaQuest is to help students learn to develop persuasive arguments supported by evidence they acquire from the course’s website or their own research. The MashpeeQuest assessment task gives students an opportunity to put these skills to use in a web-based environment that structures their work (see Box 7-1). In this example, technology plays at least two roles in enabling the assessment of complex problem solving. The first is conceptual: the information analysis skills to be assessed and the behaviors that serve as evidence are embedded within a web-based environment. The second role is more operational: since actions take place in a technological environment, some of the observations of student performance can be made
The AmericaQuest instructional program develops reasoning skills that are central to the practices of professional historians and archaeologists. The MashpeeQuest performance task is designed to tap the following subset of the skills that the AmericaQuest program is intended to foster:

Information Analysis Skills
- Ability to analyze and synthesize information from a variety of sources.
- Ability to evaluate/critique both content and sources.

Problem-Solving Skills
- Ability to synthesize disparate ideas through reasoning in a problem-solving context.
- Ability to offer reasoned arguments rather than brief guesses.
- Ability to formulate creative, well-founded theories for unsolved questions in science and history.

During instruction, students participate via the Internet in an expedition with archaeologists and historians who are uncovering clues about the fate of a Native American tribe, the Anasazi, who are believed to have abandoned their magnificent cliff dwellings in large numbers between 1200 and 1300. To collect observations of students’ acquisition of the targeted skills, the MashpeeQuest assessment task engages students in deciding a court case involving recognition of another tribe, the Mashpee Wampanoags, who some believe disappeared just as the Anasazi did. A band of people claiming Wampanoag ancestry has been trying for some years to gain recognition from the federal government as a tribe that still exists. Students are asked to investigate the evidence, select websites that provide evidence to support their claim, and justify their choices based on the evidence. They are also asked to identify one place to go to find evidence that does not support their claim, and to address how their theory of what happened to the Mashpee is still justified.

SOURCE: Adapted from Mislevy et al. (2000).
automatically (e.g., number of sources used and time per source). Other observations, such as those requiring information analysis, are made by human raters.

**Enhancing the Observation-Interpretation Linkage**

*Text Analysis and Scoring*

Extended written responses are often an excellent means of determining how well someone has understood certain concepts and can express their interrelationships. In large-scale assessment contexts, the process of reading and scoring such written products can be problematic because it is so time- and labor-intensive, even after raters have been given extensive training on standardized scoring methods. Technology tools have been developed to aid in this process by automatically scoring a variety of extended written products, such as essays. Some of the most widely used tools of this type are based on a cognitive theory of semantics called latent semantic analysis (LSA) (Landauer, Foltz, and Laham, 1998). LSA involves constructing a multidimensional semantic space that expresses the meaning of words on the basis of their co-occurrences in large amounts of text. Employing mathematical techniques, LSA can be used to “locate” units of text within this space and assign values in reference to other texts. For example, LSA can be used to estimate the semantic similarity between an essay on how the heart functions and reference pieces on cardiac structure and functioning that might be drawn from a high school text and a medical reference text.

LSA can be applied to the scoring of essays for assessment purposes in several ways. It can be used to compare a student’s essay with a set of pregraded essays at varying quality levels or with one or more model essays written by experts. Evaluation studies suggest that scores obtained from LSA systems are as reliable as those produced by pairs of human raters (Landauer, 1998). One of the benefits of such an automated approach to evaluating text is that it can provide not just a single overall score, but multiple scores on matches against different reference texts or based on subsets of the total text. These multiple scores can be useful for diagnostic purposes, as discussed subsequently.

Questions exist about public acceptance of the machine scoring of essays for high-stakes testing. There are also potential concerns about the impact of these approaches on the writing skills teachers emphasize, as well as the potential to reduce the opportunities for teacher professional development (Bennett, 1999).
Analysis of Complex Solution Strategies

We have already mentioned the possibility of using technology to assess solution moves in the context of problem-solving simulations. One of the most sophisticated examples of such analyses is offered by the IMMEX (Interactive Multimedia Exercises) program (Vendlinski and Stevens, 2000), which uses complex neural network technology to make sense of (interpret) the actions students take during the course of problem solving. IMMEX was originally developed for use in teaching and assessing the diagnostic skills of medical students. It now consists of a variety of software tools for authoring complex, multimove problem-solving tasks and for collecting performance data on those tasks, with accompanying analysis methods. The moves an individual makes in solving a problem in the IMMEX system are tracked, and the path through the solution space can be presented graphically, as well as compared against patterns previously exhibited by both skilled and less-skilled problem solvers. The IMMEX tools have been used for the design and analysis of complex problem solving in a variety of contexts, ranging from medical school to science at the college and K-12 level.

In one IMMEX problem set called True Roots, learners play the part of forensic scientists trying to identify the real parents of a baby who may have been switched with another in a maternity ward. Students can access data from various experts, such as police and hospital staff, and can conduct laboratory tests such as blood typing and DNA analysis. Students can also analyze maps of their own problem-solving patterns, in which various nodes and links represent different paths of reasoning (Lawton, 1998).

A core technology used for data analysis in the IMMEX system is artificial neural networks. These networks are used to abstract identifiable patterns of moves on a given problem from the data for many individuals who have attempted solutions, including individuals separately rated as excellent, average, or poor problem solvers. In this way, profiles can be abstracted that support the assignment of scores reflecting the accuracy and quality of the solution process. In one example of such an application, the neural network analysis of solution patterns was capable of identifying different levels of performance as defined by scores on the National Board of Medical Examiners computer-based clinical scenario exam (Casillas, Clyman, Fan, and Stevens, 2000). Similar work has been done using artificial neural network analysis tools to examine solution strategy patterns for chemistry problems (Vendlinski and Stevens, 2000).

Enhancing the Overall Design Process

The above discussion illustrates specific ways in which technology can assist in assessment design by supporting particular sets of linkages within
the assessment triangle. The Educational Testing Service has developed a more general software system, Portal, to foster cognitively based assessment design. The DISC assessment for licensing dental hygienists, described earlier, is being designed using the Portal approach.

Portal includes three interrelated components: a conceptual framework, an object model, and supporting tools. The conceptual framework is Mislevy and colleagues’ evidentiary reasoning framework (Mislevy et al., 2000), which the committee adapted and simplified to create the assessment triangle referred to throughout this report. This conceptual framework (like the assessment triangle) is at a level of generality that supports a broad range of assessment types. The object model is a set of specifications or blueprints for creating assessment design “objects.” A key idea behind Portal is that different kinds of objects are not defined for different kinds of tests; rather, the same general kinds of objects are tailored and assembled in different ways to meet different purposes. The object model for a particular assessment describes the nature of the objects and their interconnections to ensure that the design reflects attention to all the assessment elements—cognition, observation, and interpretation—and coordinates their interactions to produce a coherent assessment. Finally, the supporting tools are software tools for creating, manipulating, and coordinating a structured database that contains the elements of an assessment design.

The Portal system is a serious attempt to impose organization on the process of assessment design in a manner consistent with the thinking described elsewhere in this report on the general process of assessment design and the specific roles played by cognitive and measurement theory. Specifications and blueprints are becoming increasingly important in assessment design because, as companies and agencies develop technology-based assessments, having common yet flexible standards and language is essential for building inter-operable assessment components and processes.

STRENGTHENING THE COGNITIVE COHERENCE AMONG CURRICULUM, INSTRUCTION, AND ASSESSMENT

Technology is changing the nature of the economy and the workplace, as well as other aspects of society. To perform competently in an information society, people must be able to communicate, think, and reason effectively; solve complex problems; work with multidimensional data and sophisticated representations; navigate through a sea of information that may or may not be accurate; collaborate in diverse teams; and demonstrate self-motivation and other skills (Dede, 2000; NRC, 1999c; Secretary’s Commission on Achieving Necessary Skills [SCANS], 1991). To be prepared for this future, students must acquire different kinds of content knowledge and think-
ing skills than were emphasized in the past, and at higher levels of attainment. Numerous reports and standards documents reinforce this point of view with regard to expectations about the learning and understanding of various aspects of science, mathematics, and technology (National Council of Teachers of Mathematics, 2000; NRC, 1996, 1999b; SCANS, 1991).

Consistent with these trends and pressures, many schools are moving toward instructional methods that encourage students to learn mathematics and science content in greater depth, and to learn advanced thinking skills through longer-term projects that emphasize a scientific process of inquiry, involve the use of complex materials, and cut across multiple disciplines. Many of these reforms incorporate findings from cognitive research, and many assign technology a central role in this transformation.

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, more timely, and more seamlessly interwoven with curriculum and instruction.

In this section we consider two ways in which this integration is under way. In the first set of scenarios, the focus is on using technology tools and systems to assist in the integration of cognitively based assessment into classroom practice in ways consistent with the discussion in Chapter 6. In the second set of scenarios, the focus is on technology-enhanced learning environments that provide for a more thorough integration of curriculum, instruction, and assessment guided by models of cognition and learning.

**Facilitating Formative Assessment**

As discussed in earlier chapters, the most useful kinds of assessment for enhancing student learning often 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. In diagnostic assessments of individual learning, for example, significant amounts of information must be collected, interpreted, and reported. No individual, 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 difficult or impractical for a classroom teacher.
Several examples illustrate how technology can help infuse ongoing formative assessment into the learning process. Notable among them are intelligent tutors, described previously in this report (see Chapters 3 and 6). Computer technology facilitates fine-grained analysis of the learner's cognitive processes and knowledge states in terms of the theoretical models of domain learning embedded within a tutoring system. Without technology, it would not be possible to provide individualized and interactive instruction, extract key features of the learner's responses, immediately analyze student errors, and offer relevant feedback for remediating those errors—let alone do so in a way that would be barely identifiable as assessment to the learner.

Another example is the DIAGNOSER computerized assessment tool used in the Facets-based instructional program (described in Chapters 5 and 6). DIAGNOSER allows students to test their own understanding as they work through various modules on key math and science concepts (Minstrell, Stimpson, and Hunt, 1992). On the basis of students' responses to the carefully constructed questions, the program can pinpoint areas of possible misunderstanding, give feedback on reasoning strategies, and prescribe relevant instruction. The program also keeps track of student responses, which teachers can use for monitoring overall class performance. As illustrated in Chapter 5, the assessment questions are in a multiple-choice format, in which each possible answer corresponds to a facet of reasoning. Although the assessment format is rather conventional, this basic model could be extended to other, more complex applications. Minstrell (2000; see also Hunt and Minstrell, 1994) has reported data on how use of a Facets instructional approach that includes the DIAGNOSER software system has significantly enhanced levels of student learning in high school physics.

ThinkerTools (also described in Chapter 6) is a computer-enhanced middle school science curriculum that promotes metacognitive skills by encouraging students to evaluate their own and each others' work using a set of well-considered criteria (White and Frederiksen, 2000). Students propose and test competing theories, carry out experiments using the computer and real-world materials, compare their findings, and try to reach consensus about the best model. The software enables students to simulate experiments, such as turning gravity on and off, that would be impossible to perform in the real world, and to accurately measure distances, times, and velocities that would similarly be difficult to measure in live experiments. Feedback about the correctness of student conjectures is provided as part of the system. According to a project evaluation, students who learned to use the self-assessment criteria produced higher-quality projects than those who did not, and the benefits were particularly obvious for lower-achieving students.

The IMMEX program was discussed in the preceding section as an example of a powerful set of technology tools for assessing student problem
solving in areas of science. Teachers participate in professional development experiences with the IMMEX system and learn to use the various tools to develop problem sets that then become part of their overall curriculum-instruction-assessment environment. Students work on problems in the IMMEX system, and feedback can be generated at multiple levels of detail for both teacher and student. For the teacher, data are available on how individual students and whole classes are doing on particular problems, as well as over time on multiple problems. At a deeper level of analysis, teachers and students can obtain visual maps of their search through the problem space for the solution to a given problem. These maps are rich in information, and teachers and students can use them in multiple ways to review and discuss the problem-solving process. By comparing earlier maps with later ones, teachers and students can also judge refinements in problem-solving processes and strategies (see Vendelinski and Stevens, 2000, for examples of this process).

Technology-based assessment tools are not limited to mathematics and science. Summary Street, experimental software for language arts, helps middle school students improve their reading comprehension and writing skills by asking them to write summaries of materials they have read. Using a text analysis program based on LSA, the computer compares the summary with the original text and analyzes it for certain information and features. The program also gives students feedback on how to improve their summaries before showing them to their teachers (Kintsch, Steinhart, Stahl, LSA Research Group, Matthews, and Lamb, 2000). Research with this system has shown substantial improvements in students’ summary generation skills, which generalize to other classes and are independent of having further access to the Summary Street software program (Kintsch et al., 2000).

The preceding examples are not an exhaustive list of instances in which technology has been used to create formative assessment tools that incorporate various aspects of cognitive and measurement theory. An especially important additional example is the integration of concept mapping tools, discussed earlier, into instructional activities (see Mintzes et al., 1998). In the next section, we consider other examples of the use of technology-assisted formative assessment tools in the instructional process. In many of these cases, the tools are an integral part of a more comprehensive, technology-enhanced learning environment.

**Technology-Enhanced Learning Environments**

Some of the most powerful technology-enhanced innovations that link curriculum, instruction, and assessment focus on aspects of the mathematics and science curriculum that have heretofore been difficult to teach. Many of these designs were developed jointly by researchers and educators using
findings from cognitive science and classroom practice. Students and teachers in these environments use technology to conduct research, solve problems, analyze data, interact with others, track progress, present their results, and accomplish other goals. Typically, these environments emphasize learning through the processes of inquiry and collaborative, problem-based learning, and their learning goals are generally consistent with those discussed in the National Council of Teachers of Mathematics (2000) mathematics standards and the NRC (1996) science standards. In many ways, they also go beyond current standards by emphasizing new learning outcomes students need to master to perform competently in an information society.

In these environments, it is not uncommon for learners to form live and on-line communities of practice, evaluating their own and each others’ reasoning, hypotheses, and work products. These programs also tend to have multiple learning goals; in addition to teaching students important concepts in biology, physics, or earth science, for example, they may seek to teach students to think, work, and communicate as scientists do. These environments, examples of which are described below, illustrate the importance of carefully designed assessment to an effective learning environment (NRC, 1999d) and show how technology makes possible the integration of assessment with instruction in powerful ways.

Use of Technology to Enhance Learning Environments: Examples

SMART Model An example of embedding assessment strategies within extended-inquiry activities can be found in work pursued by the Cognition and Technology Group at Vanderbilt University (CTGV) on the development of a conceptual model for integrating curriculum, instruction, and assessment in science and mathematics (Barron et al., 1995; CTGV, 1994, 1997). The resultant SMART (Scientific and Mathematical Arenas for Refining Thinking) Model incorporates frequent opportunities for formative assessment by both students and teachers, and reflects an emphasis on self-assessment to help students develop the ability to monitor their own understanding and find resources to deepen it when necessary (Vye et al., 1998). The SMART Model involves the explicit design of multiple cycles of problem solving, self-assessment, and revision in an overall problem-based to project-based learning environment.

Activity in the problem-based learning portion of SMART typically begins with a video problem scenario, for example, from the Adventures of Jasper Woodbury mathematics problem-solving series (CTGV, 1997) or the Scientists in Action series. An example of the latter is the Stones River Mystery (Sherwood, Petrosino, Lin, and CTGV, 1998), which tells the story of a group of high school students who, in collaboration with a biologist and a
hydrologist, are monitoring the water in Stones River. The video shows the team visiting the river and conducting various water quality tests. Students in the classroom are asked to assess the water quality at a second site on the river. They are challenged to select tools they can use to sample macroinvertebrates and test dissolved oxygen, to conduct these tests, and to interpret the data relative to previous data from the same site. Ultimately, they find that the river is polluted as a result of illegal dumping of restaurant grease. Students must then decide how to clean up the pollution.

The problem-based learning activity includes three sequential modules: macroinvertebrate sampling, dissolved oxygen testing, and pollution cleanup. The modules are preliminary to the project-based activity, in which students conduct actual water quality testing at a local river. In executing the latter, they are provided with a set of criteria by which an external agency will evaluate written reports and accompanying videotaped presentations.

The ability of students and teachers to progress through the various cycles of work and revision within each module and devise an effective solution to the larger problem depends on a variety of resource materials carefully designed to assist in the learning and assessment process (see Box 7-2). Students who use these resources and tools learn significantly more than students who go through the same instructional sequence for the same amount of time, but without the benefit of the tools and the embedded formative assessment activities. Furthermore, their performance in a related project-based learning activity is significantly enhanced (Barron et al., 1995).

Genscope™ This is an innovative computer-based program designed to help students learn key concepts of genetics and develop scientific reasoning skills (see Hickey, Kindfield, and Horwitz, 1999; Horwitz, 1998). The program includes curriculum, instructional components, and assessments. The centerpiece is an open-ended software tool that permits students to manipulate models of genetic information at multiple levels, including cells, family trees, and whole populations. Using Genscope™, students can create and vary the biological traits of an imaginary species of dragons—for example, by altering a gene that codes for the dragon’s color and exploring how this alteration affects generations of offspring and the survivability of a population.

The developers of Genscope™ have pursued various approaches to the assessment of student learning outcomes. To compare the performance of ninth graders who used this program and those in more traditional classrooms, the researchers administered a paper-and-pencil test. They concluded that many Genscope™ students were not developing higher-order reasoning skills as intended. They also found that some classrooms that used these curriculum materials did not complete the computerized activities because of various logistical problems; nevertheless, the students in these classrooms...
Solution of the Stones River Mystery requires students to work through three successive activity modules focused on macroinvertebrate sampling, dissolved oxygen testing, and pollution cleanup. Each module follows the same cycle of activities: initial selection of a method for testing or cleanup, feedback on the initial choice, revision of the choice, and a culminating task. Within each activity module, selection, feedback, and revision make use of the SMART web site, which organizes the overall process and supports three high-level functions.

First, it provides individualized feedback to students and serves as a formative evaluation tool. As with DIAGNOSER, the feedback suggests aspects of students’ work that are in need of revision and classroom resources students can use to help them make these revisions. The feedback does not tell students the “right answer.” Instead, it sets a course for independent inquiry by the student. The Web feedback is generated from data entered by individual students.

The second function of SMART web site is to collect, organize, and display the data collected from multiple distributed classrooms, a function performed by SMART Lab. Data displays are automatically updated as students submit new data. The data in SMART Lab consist of students’ answers to problems and explanations for their answers. Data from each class can be displayed separately from those of the distributed classroom. This feature enables the teacher and her or his class to discuss different solution strategies, and in the process address important concepts and misconceptions. These discussions provide a rich source of information for the teacher on how students are thinking about a problem and are designed to stimulate further student reflection.

The third function of SMART web site is performed by Kids Online. Students are presented with the explanations of student-actors. The explanations are text-based with audio narration, and they are errorful by design. Students are asked to critically evaluate the explanations and provide feedback to the student-actors. The errors seed thinking and discussion on concepts that are frequently misconceived by students. At the same time, students learn important critical evaluation skills.
showed the same reasoning gains as their schoolmates who used the software. To address the first concern, the researchers developed a set of formative assessments, including both worksheets and computer activities, that encouraged students to practice specific reasoning skills. To address the second concern, the developers refined the program and carefully designed a follow-up evaluation with more rigorous controls and comparison groups; this evaluation showed that GenScope™ was notably more effective than traditional methods in improving students’ reasoning abilities in genetics.

Recently, the researchers have begun testing a new assessment software tool, BioLogica, which embeds formative assessment into the computerized GenScope™ activities. This system poses sequences of challenges to students as they work, monitors their actions, intervenes with hints or feedback, asks questions intended to elicit student understanding, provides tools the students can use to meet the challenge, and directs them to summon the teacher for discussion. The BioLogica scripts are based on numerous hours of observation and questioning of children as they worked. The system can also create personal portfolios of a student’s notes and images and record the ongoing interactions of multiple students in a massive log that can be used for formative or summative assessment.

**Knowledge Integration Environment**  The Knowledge Integration Environment (KIE) is another technology-based instructional environment that engages middle and high school students in scientific inquiry using the Internet. KIE consists of a set of complementary software components that provide browsing, note-taking, discussion, argument-building, and guidance capabilities. The instructional goal is to foster knowledge integration by encouraging students to make connections between scientific concepts and relate these concepts to personally relevant situations and problems. As part of the design process, developers conducted research with students in real classroom environments to gain a better understanding of the cognitive benefits of different kinds of prompts, ways in which perspective taking can be scaffolded, and the effects of evidence presentation on student interpretation (Bell, 1997). The KIE curriculum consists of units, called “projects,” that typically last three to ten class periods. Projects include debates, critiques, and design projects. Box 7-3 describes an example of a KIE project.

KIE is related to another program, the Computer as Learning Partner (CLP), which is based on many of the same principles. The developers of CLP have used a variety of measures to evaluate its impact on learning (Linn and Hsi, 2000). Students engaged in a CLP unit on heat and temperature, for

---

1Scaffolding consists of explicit and sequentially organized support and guidance about possible strategies.
example, showed substantial gains in understanding and outperformed comparable twelfth-grade students in traditional classrooms. Researchers are also working on designing assessments that can similarly be used to evaluate the impact of KIE on learning. Early results from this research suggest benefits comparable to those of CLP (Linn and Hsi, 2000).

**Additional Examples**  The list of technology-enhanced learning environments, as well as other applications of technology tools in educational settings, continues to grow (see CILT.org and LETUS.org). Two other environments with a significant research and development history bear mentioning for their consideration of issues associated with the integration of curriculum, instruction, and assessment. The first is the CoVis (Learning Through Collaborative Visualization) project. The goal of the CoVis project is to support the formation and work of learning communities by providing media-rich communication and scientific visualization tools in a highly interactive, networked, collaborative context. For instance, students collaborate with distant peers to study meteorological phenomena using computational tools. A variety of technological tools make it possible for participants to record, carry out, and discuss their project work with peers in other locations (Edelson, 1997).

The second such environment is MOOSE Crossing, an innovative technology-based program that uses an on-line learning environment to improve the reading, writing, and programming skills of children aged 8 to 14. The program enables children to build virtual objects, creatures, and places and to program their creations to move, change, and interact in the virtual environment (Bruckman, 1998). The broader goal of the project is to create a self-directed and self-supporting on-line community of learners who provide technical and social support, collaborate on activities, and share examples of completed work. Initially the project encompassed no formal assessment, consistent with its philosophy of encouraging self-motivated learning, but over time the developers realized the need to evaluate what students were learning. Assessments of programming skills indicated that most participants were not attempting to write programming scripts of any complexity, so the project designers introduced a merit badge system to serve as a motivator, scaffold, and assessment tool. To win a badge, students must work with a mentor to complete a portfolio, which is reviewed by anonymous reviewers.

**Assessment Issues and Challenges for Technology-Enhanced Learning Environments**

As the preceding examples illustrate, many technology-enhanced learning environments have integrated formative and summative assessments into
KIE is organized around projects that involve students in using the Internet. The How Far Does Light Go? project (Bell, 1997) asks students to contrast two theoretical positions about the propagation of light using text and multimedia evidence derived from both scientific and everyday sources. The first theoretical position in the debate is the scientifically normative view that “light goes forever until it is absorbed,” while the second is the more phenomenological perception that “light dies out as you move farther from a light source.”

Students begin the project by stating their personal position on how far light goes. They review a set of evidence and determine where each piece fits into the debate. After creating some evidence of their own, the students then synthesize the evidence by selecting the pieces that in their opinion factor most prominently into the debate and composing written explanations to that effect. The result is a scientific argument supporting one of the two theoretical positions. Student teams present their arguments in a classroom discussion and respond to questions from the other students and the teacher. As the project concludes, students are asked to reflect upon issues that arose during the activity and once again state their position in the debate.

Sensemaker is one software component of KIE. It provides a spatial and categorical representation for a collection of web-based evidence. The sample screen below shows a SenseMaker argument constructed jointly by a student pair for use in a classroom debate as part of the How Far Does Light Go? project.

Several factors led to this realization. First, after analyzing students’ interactions, researchers and teachers often discovered that many learners re-
quired more scaffolding to motivate them to tackle challenging tasks and help them acquire a deep level of understanding. Thus, components were added to programs such as GenScope™, Moose Crossing, and SMART to give students more advice, encouragement, and practice.

Second, researchers found that teachers and students needed formative assessment to help them monitor what was being learned and develop their metacognitive skills (CTGV, 1997; White and Frederiksen, 1998). Consequently, many of these environments included methods of recording and analyzing students' inquiry processes and ways of encouraging them to reflect on and

revise their thinking. An interesting example is the cognitively based scheme for computerized diagnosis of study skills (e.g., self-explanation) recently produced and tested by Conati and VanLehn (1999). The development of metacognitive skills is also an explicit part of the designs used in SMART.

Third, researchers and educators realized that to document the learning effects of these innovations for parents, policy makers, funding agencies, and other outside audiences, it would be necessary to have assessments that captured the complex knowledge and skills that inquiry-based learning environments are designed to foster. Not surprisingly, traditional assessments of mathematics and science typically reveal little about the benefits of these kinds of learning environments. Indeed, one can understand why there is often no evidence of benefit when typical standardized tests are used to evaluate the learning effects of many technology-based instructional programs. The use of such tests constitutes an instance of a poor fit between the observation and cognitive elements of the assessment triangle. The tasks used for typical standardized tests provide observations that align with a student model focused on specific types of declarative and procedural knowledge that may or may not have been acquired with the assistance of the technology-based programs. Thus, it should come as no surprise that there is often a perceived mismatch between the learning goals of many educational technology programs and the data obtained from standardized tests. Despite their inappropriateness, however, many persist in using such data as the primary basis for judging the effectiveness and value of investments in educational technology.

Unfortunately, this situation poses a significant assessment and evaluation challenge for the designers and implementers of technology-enhanced learning environments. For example, if such environments are to be implemented on a wider scale, evidence must be produced that students are learning things of value, and this evidence must be convincing and accepted as valid by outside audiences. In many technology-enhanced learning environments, the data provided are from assessments that are highly contextualized: assessment observations are made while students are engaged in learning activities, and the model used to interpret these observations is linked specifically to that project. Other concerns relate to the technical quality of the assessment information. Is it derived from a representative sample of learners? Are the results generalizable to the broad learning goals in that domain? Are the data objective and technically defensible? Such concerns often make it difficult to use assessment data closely tied to the learning environment to convince educators and the public of the value of these new kinds of learning environments. Without such data it is difficult to expand the audience for these programs so that they are used on a larger scale. This dilemma represents yet another example of the point, made earlier in
this report, that assessment that serves the purpose of supporting student learning may not serve the purpose of program evaluation equally well.

While this dilemma is complex and often poorly understood, it can begin to be addressed by starting with a clear definition of both the goals for learning in such environments and the targets of inference. By following the design principles set forth in this report, it is possible to design fair assessments of student attainment that are not totally embedded in the learning environment or confounded with technology use. Assessing the knowledge students acquire in specific technology-enhanced learning environments requires tasks and observations designed to provide evidence consistent with an appropriate student model. The latter identifies the specific knowledge and skills students are expected to learn and the precise form of that knowledge, including what aspects are tied to specific technology tools. An interesting example of this principled approach to assessment design is the Mashpee Quest task (described earlier in Box 7-1) (Mislevy et al., 2000).

LINKAGE OF ASSESSMENTS FOR CLASSROOM LEARNING AND ACCOUNTABILITY PURPOSES

A Vision of the Possible

While it is always risky to predict the future, it appears clear that advances in technology will continue to impact the world of education in powerful and provocative ways. 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 a more individualized and mastery-oriented approach to learning. 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, and by whom.

One of the limiting factors in effectively integrating assessment into educational systems to address the range of questions that need to be answered about student achievement is the lack of models of student learning for many aspects of the curriculum. This situation will change over time, and it will become possible to incorporate much of the necessary theoretical and empirical knowledge into technology-based systems for instruction and assessment.

It is both intriguing and useful to consider the possibilities that might arise if assessment were integrated into instruction in multiple curricular areas, and the resultant information about student accomplishment and understanding were collected with the aid of technology. In such a world,
programs of on-demand external assessment might not be necessary. 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.

Extensive technology-based systems that link curriculum, instruction, and assessment at the classroom level might enable a shift from today’s assessment systems that use different kinds of assessments for different purposes to a balanced design in which the features of comprehensiveness, coherence, and continuity would be assured (see Chapter 6). One can imagine a future in which the audit function of external assessments would be significantly reduced or even unnecessary because the information needed to assess students at the levels of description appropriate for various external assessment purposes could be derived from the data streams generated by students in and out of their classrooms. Technology could offer ways of creating over time a complex stream of data about how students think and reason while engaged in important learning activities. Information for assessment purposes could be extracted from this stream and used to serve both classroom and external assessment needs, including providing individual feedback to students for reflection about their metacognitive habits. To realize this vision, research on the data representations and analysis methods best suited for different audiences and different assessment objectives would clearly be needed.

A metaphor for this shift 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 checkout and barcodes for all items, these businesses 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 no longer have to interrupt the normal instructional process at various times during the year to administer external tests to students.

While the committee is divided as to the practicality and advisability of pursuing the scenario just described, we offer it as food for thought about future states that might be imagined or invented. Regardless of how far one wishes to carry such a vision, it is clear that technological advances will allow for the attainment of many of the goals for assessment envisioned in this report. When powerful technology-based systems are implemented in multiple classrooms, rich sources of information about student learning will be continuously available across wide segments of the curriculum and for individual learners over extended periods of time. The major issue is not whether this type of data collection and information analysis is feasible in
the future. Rather, the issue is how the world of education will anticipate and embrace 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 aid in student learning.

It is sometimes noted that the best way to predict the future is to invent it. Multiple futures for educational assessment could be invented on the basis of synergies that exist among information technologies and contemporary knowledge of cognition and measurement. In considering these futures, however, one must also explore a number of associated issues and challenges.

Issues and Challenges

Visions of assessment integrated with instruction and of the availability of complex forms of data coexist with other visions of education’s future as a process of “distributed learning”: educational activities orchestrated by means of information technology across classrooms, workplaces, homes, and community settings and based on a mix of presentational and constructivist pedagogies (guided inquiry, collaborative learning, mentoring) (Dede, 2000). Recent advances in groupware and experiential simulation enable guided, collaborative inquiry-based learning even though students are in different locations and often are not on line at the same time. With the aid of telementors, students can create, share, and master knowledge about authentic real-world problems. Through a mix of emerging instructional media, learners and educators can engage in synchronous or asynchronous interaction: face-to-face or in disembodied fashion or as an “avatar” expressing an alternate form of individual identity. Instruction can be distributed across space, time, and multiple interactive media. These uses of technology for distributed learning add a further layer of complexity to issues raised by the potential for using technology to achieve the integrated forms of assessment envisioned in this report.

Policy Issues

Although powerful, distributed learning strategies render assessment and evaluation for comparative and longitudinal purposes potentially more problematic. A major question is whether assessment strategies connected to such environments would be accepted by policy makers and others who make key decisions about current investments in large-scale assessment programs. Many people will be unfamiliar with these approaches and concerned about their fairness, as well as their fit with existing standards in some states. Questions also arise about how to compare the performance of
students in technology-enhanced environments with that of students in more traditional classrooms, as well as students who are home-schooled or educated in settings without these technologies.

**Pragmatic Issues**

Issues of utility, practicality, and cost would have to be addressed to realize the vision of integrated assessments set forth above. First is the question of how different users would make sense of the masses of data that might be available and which kinds of information should be extracted for various purposes. Technical issues of data quality and validity would also become paramount. With regard to costs, designing integrated assessments would be labor-intensive and expensive, but technological tools that would aid in the design and implementation process could eventually reduce the costs and effort involved. The costs would also need to be spread over time and amortized by generating information that could be used for multiple purposes: diagnosis and instructional feedback, classroom summative assessment, and accountability.

**Equity Issues**

Many of the innovations described in this chapter have the potential to improve educational equity. Initial evaluations of technology-enhanced learning projects have suggested that they may have the power to increase learning among students in low-income areas and those with disabilities (Koedinger, Anderson, Hadley, and Mark, 1997; White and Frederiksen, 1998). Technology-enhanced instruction and assessment may also be effective with students who have various learning styles—for example, students who learn better with visual supports than with text or who learn better through movement and hands-on activities. These approaches can also encourage communication and performances by students who are silent and passive in classroom settings, are not native English speakers, or are insecure about their capabilities. Technology is already being used to assess students with physical disabilities and other learners whose special needs preclude representative performance using traditional media for measurement (Dede, 2000).

Evidence indicates that hands-on methods may be better than textbook-oriented instruction for promoting science mastery among students with learning disabilities. At-risk students may make proportionately greater gains than other students with technology-enhanced programs. The GenScope™ program, for example, was conducted in three ninth-grade biology classrooms of students who were not college bound; about half of the students in all three classrooms had been identified as having learning or behavioral disabilities. An auspicious finding of the program evaluation was that the
gains in reasoning skills in the two classrooms using GenScope™ were much greater than those in the classroom following a more traditional approach (Hickey et al., 1999).

An example of the potential of technology-enhanced learning can be found in the Union City, New Jersey, school district, which has reshaped its curriculum, instruction, and assessment around technology. Evaluations have demonstrated positive and impressive gains in student learning in this diverse, underfinanced district (Chang et al., 1998).

Privacy Issues

Yet another set of critical issues relates to privacy. On the one hand, an integrated approach to assessment would eliminate the need for large banks of items to maintain test security, because the individual trail of actions taken while working on a problem would form the items. On the other hand, such an approach raises different kinds of security issues. When assessments record students’ actions as they work, a tension exists between the need to protect students’ privacy and the need to obtain information for demonstrating program effectiveness. Questions arise of what consequences these new forms of embedded assessment would have for learning and experimentation if students (and teachers) knew that all or many of their actions—successes and failures—were being recorded. Other questions relate to how much information would be sampled and whether that information would remain private.

Learners and parents have a right to know how their performance will be judged and how the data being collected will be used. Existing projects that record student interactions for external evaluations have obtained informed consent from parents to extract certain kinds of information. Scaling up these kinds of assessments to collect and extract information from complex data streams could be viewed as considerably more invasive and could escalate concerns about student privacy.

CONCLUSIONS

Information technologies are helping to remove some of the constraints that have limited assessment practice in the past. Assessment tasks no longer need be confined to paper-and-pencil formats, and the entire burden of classroom assessment no longer need fall on the teacher. At the same time, technology will not in and of itself improve educational assessment. Improved methods of assessment require a design process that connects the three elements of the assessment triangle to ensure that the theory of cognition, the observations, and the interpretation process work together to support the intended inferences. Fortunately, there exist multiple examples of tech-
Some of the most intriguing applications of technology extend the nature of the problems that can be presented and the knowledge and cognitive processes that can be assessed. By enriching task environments through the use of multimedia, interactivity, and control over the stimulus display, it is possible to assess a much wider array of cognitive competencies than has heretofore been feasible. New capabilities enabled by technology include directly assessing problem-solving skills, making visible sequences of actions taken by learners in solving problems, and modeling and simulating complex reasoning tasks. Technology also makes possible data collection on concept organization and other aspects of students’ knowledge structures, as well as representations of their participation in discussions and group projects. A significant contribution of technology has been to the design of systems for implementing sophisticated classroom-based formative assessment practices. Technology-based systems have been developed to support individualized instruction by extracting key features of learners’ responses, analyzing patterns of correct and incorrect reasoning, and providing rapid and informative feedback to both student and teacher.

A major change in education has resulted from the influence of technology on what is taught and how. Schools are placing more emphasis on teaching critical content in greater depth. Examples include the teaching of advanced thinking and reasoning skills within a discipline through the use of technology-mediated projects involving long-term inquiry. Such projects often integrate content and learning across disciplines, as well as integrate assessment with curriculum and instruction in powerful ways.

A possibility for the future arises from the projected growth across curricular areas of technology-based assessment embedded in instructional settings. Increased availability of such systems could make it possible to pursue balanced designs representing a more coordinated and coherent assessment system. Information from such assessments could possibly be used for multiple purposes, including the audit function associated with many existing external assessments.

Finally, technology holds great promise for enhancing educational assessment at multiple levels of practice, but its use for this purpose also raises issues of utility, practicality, cost, equity, and privacy. These issues will need to be addressed as technology applications in education and assessment continue to expand, evolve, and converge.