Part I

Overview and Background
Rethinking the Foundations of Assessment

The time is right to rethink the fundamental scientific principles and philosophical assumptions that underlie current approaches to educational assessment. These approaches have been in place for decades and have served a number of purposes quite well. But the world has changed substantially since those approaches were first developed, and the foundations on which they were built may not support the newer purposes to which assessments may be put. Moreover, advances in the understanding and measurement of learning bring new assumptions into play and offer the potential for a much richer and more coherent set of assessment practices. In this volume, the Committee on the Foundations of Assessment outlines these new understandings and proposes a new approach to assessment.

CHARGE TO THE COMMITTEE

The Committee on the Foundations of Assessment was convened in January 1998 by the National Research Council (NRC) with support from the National Science Foundation. The committee’s charge was to review and synthesize advances in the cognitive sciences and to explore their implications for improving educational assessment in general and assessment of science and mathematics education in particular. The committee was also charged with evaluating the extent to which evolving assessment practices in U.S. schools were derived from research on cognition and learning, as well as helping to improve public understanding of current and emerging assessment practices and uses. The committee approached these three objectives as interconnected themes rather than as separate tasks.
SCOPE OF THE STUDY

The committee considered the implications of advances in the cognitive and measurement sciences for both classroom and large-scale assessment. Consistent with its charge, the committee focused primarily on assessment in science and mathematics education. Although new concepts of assessment could easily apply to other disciplines, science and mathematics hold particular promise for rethinking assessment because of the substantial body of important research and design work already done in these disciplines. Because science and mathematics also have a major impact on the nation’s technological and economic progress, they have been primary targets for education reform at the national and state levels, as well as a focus of concern in international comparative studies. Furthermore, there are persistent disparities among ethnic, geographic, and socioeconomic groups in access to quality K-12 science and mathematics instruction. Black, Hispanic, and Native American youth continue to lag far behind Whites and Asians in the amount of coursework taken in these subjects and in levels of achievement; this gap negatively affects their access to certain careers and workforce skills. Better assessment, curriculum, and instruction could help educators diagnose the needs of at-risk students and tailor improvements to meet those needs.

The committee also focused on the assessment of school achievement, or the outcomes of schooling, and gave less emphasis to predictive tests (such as college selection tests) that are intended to project how successful an individual will be in a future situation. We had several reasons for this emphasis. First, when one considers the use of assessments at the classroom, district, state, and national levels in any given year, it is clear that the assessment of academic achievement is far more extensive than predictive testing. Second, many advances in cognitive science have already been applied to the study and design of predictive instruments, such as assessments of aptitude or ability. Much less effort has been expended on the application of advances in the cognitive and measurement sciences to issues of assessing academic content knowledge, including the use of such information to aid teaching and learning. Finally, the committee believed that the principles and practices uncovered through a focus on the assessment of academic achievement would generally apply also to what we view as the more circumscribed case of predictive testing.

Our hope is that by reviewing advances in the sciences of how people learn and how such learning can be measured, and by suggesting steps for future research and development, this report will help lay the foundation for a significant leap forward in the field of assessment. The committee envisions a new generation of educational assessments that better serve the goal of equity. Needed are assessments that help all students learn and succeed
in school by making as clear as possible the nature of their accomplishments and the progress of their learning.

**CONTEXT**

In this first chapter we embed the discussion of classroom and large-scale assessment in a broader context by considering the social, technological, and educational setting in which it operates. The discussion of context is organized around four broad themes:

- Any assessment is based on three interconnected elements or foundations: the aspects of achievement that are to be assessed (cognition), the tasks used to collect evidence about students’ achievement (observation), and the methods used to analyze the evidence resulting from the tasks (interpretation). To understand and improve educational assessment, the principles and beliefs underlying each of these elements, as well as their interrelationships, must be made explicit.
- Recent developments in society and technology are transforming people’s ideas about the competencies students should develop. At the same time, education policy makers are attempting to respond to many of the societal changes by redefining what all students should learn. These trends have profound implications for assessment.
- Existing assessments are the product of prior theories of learning and measurement. While adherence to these theories has contributed to the enduring strengths of these assessments, it has also contributed to some of their limitations and impeded progress in assessment design.
- Alternative conceptions of learning and measurement now exist that offer the possibility to establish new foundations for enhanced assessment practices that can better support learning.

The following subsections elaborate on each of these themes in turn. Some of the key terms used in the discussion and throughout this report are defined in Box 1-1.

**The Significance of Foundations**

From teachers’ informal quizzes to nationally administered standardized tests, assessments have long been an integral part of the educational process. Educational assessments assist teachers, students, and parents in determining how well students are learning. They help teachers understand how to adapt instruction on the basis of evidence of student learning. They help principals and superintendents document the progress of individual stu-
The foundations influence all aspects of an assessment’s design and use, including content, format, scoring, reporting, and use of the results. Even though these fundamental principles are sometimes more implicit than explicit, they are still influential. In fact, it is often the tacit nature of the found-
dations and the failure to question basic assumptions that creates conflicts about the meaning and value of assessment results.

Advances in the study of thinking and learning (cognitive science) and in the field of measurement (psychometrics) have stimulated people to think in new ways about how students learn and what they know, what is therefore worth assessing, and how to obtain useful information about student competencies. Numerous researchers interested in problems of educational assessment have argued that, if brought together, advances in the cognitive and measurement sciences could provide a powerful basis for refashioning educational assessment (e.g., Baker, 1997; Glaser and Silver, 1994; Messick, 1984; Mislevy, 1994; National Academy of Education, 1996; Nichols, 1994; National Research Council [NRC], 1999b; Pellegrino, Baxter, and Glaser, 1999; Snow and Lohman, 1989; Wilson and Adams, 1996). Indeed, the merger could be mutually beneficial, with the potential to catalyze further advances in both fields.

Such developments, if vigorously pursued, could have significant long-term implications for the field of assessment and for education in general. Unfortunately, the theoretical foundations of assessment seldom receive explicit attention during most discussions about testing policy and practice. Short-term issues of implementation, test use, or score interpretation tend to take precedence, especially in the context of many large-scale testing programs (NRC, 1999b). It is interesting to note, however, that some of today’s most pressing issues, such as whether current assessments for accountability encourage effective teaching and learning, ultimately rest on an analysis of the fundamental beliefs about how people learn and how to measure such learning that underlie current practices. For many reasons, the present climate offers an opportune time to rethink these theoretical underpinnings of assessment, particularly in an atmosphere, such as that surrounding the committee’s deliberations, not charged with the polarities and politics that often envelop discussions of the technical merits of specific testing programs and practices.

**Changing Expectations for Learning**

Major societal, economic, and technological changes have transformed public conceptions about the kinds of knowledge and skills schools should teach and assessments should measure (Secretary’s Commission on Achieving Necessary Skills, 1991). These developments have sparked widespread debate and activity in the field of assessment. The efforts under way in every state to reform education policy and practice through the implementation of higher standards for students and teachers have focused to a large extent on assessment, resulting in a major increase in the amount of testing and in the emphasis placed on its results (Education Week, 1999). The following sub-
sections briefly review these trends, which are changing expectations for
student learning and the assessment of that learning.

Societal, Economic, and Technological Changes

Societal, economic, and technological changes are transforming the world
of work. The workforce is becoming more diverse, boundaries between jobs
are blurring, and work is being structured in more varying ways (NRC, 1999a).
This restructuring often increases the skills workers need to do their jobs.
For example, many manufacturing plants are introducing sophisticated in-
formation technologies and training employees to participate in work teams
(Appelbaum, Bailey, Berg, and Kalleberg, 2000). Reflecting these transfor-
mations in work, jobs requiring specialized skills and postsecondary educa-
tion are expected to grow more quickly than other types of jobs in the
coming years (Bureau of Labor Statistics, 2000).

To succeed in this increasingly competitive economy, all students, not
just a few, must learn how to communicate, to think and reason effectively,
to solve complex problems, to work with multidimensional data and sophis-
ticated representations, to make judgments about the accuracy of masses of
information, to collaborate in diverse teams, and to demonstrate self-motiva-
tion (Barley and Orr, 1997; NRC, 1999a, 2001). As the U.S. economy contin-
ues its transformation from manufacturing to services and, within services,
to an “information economy,” many more jobs are requiring higher-level
skills than in the past. Many routine tasks are now automated through the
use of information technology, decreasing the demand for workers to per-
form them. Conversely, the demand for workers with high-level cognitive
skills has grown as a result of the increased use of information technology in
the workplace (Bresnahan, Brynjolfsson, and Hitt, 1999). For example, orga-
nizations have become dependent upon quick e-mail interactions instead of
slow iterations of memoranda and replies. Individuals not prepared to be
quickly but effectively reflective are at a disadvantage in such an environ-
ment.

Technology is also influencing curriculum, changing what and how stu-
dents are learning, with implications for the types of competencies that should
be assessed. New information and communications technologies present
students with opportunities to apply complex content and skills that are
difficult to tap through traditional instruction. In the Weather Visualizer pro-
gram, for example, students use sophisticated computer tools to observe
complex weather data and construct their own weather forecasts (Edelson,
Gordon, and Pea, 1999).

These changes mean that more is being demanded of all aspects of
education, including assessment. Assessments must tap a broader range of
competencies than in the past. They must capture the more complex skills
and deeper content knowledge reflected in new expectations for learning. They must accurately measure higher levels of achievement while also providing meaningful information about students who still perform below expectations. All of these trends are being played out on a large scale in the drive to set challenging standards for student learning.

**An Era of Higher Standards and High-Stakes Tests**

Assessment has been greatly influenced by the movement during the past two decades aimed at raising educational quality by setting challenging academic standards. At the national level, professional associations of subject matter specialists have developed widely disseminated standards outlining the content knowledge, skills, and procedures schools should teach in mathematics, science, and other areas. These efforts include, among others, the mathematics standards developed by the National Council of Teachers of Mathematics (2000), the science standards developed by the NRC (1996), and the standards in several subjects developed by New Standards (e.g., New Standards™, 1997), a privately funded organization.

In addition, virtually every state and many large school districts have standards in place outlining what all students should know and be able to do in core subjects. These standards are intended to guide both practice and policy at the state and district levels, including the development of large-scale assessments of student performance. The process of developing and implementing standards at the national and local levels has advanced public dialogue and furthered professional consensus about the kinds of knowledge and skills that are important for students to learn at various stages of their education. Many of the standards developed by states, school districts, and professional groups emphasize that it is important for students not only to attain a deep understanding of the content of various subjects, but also to develop the sophisticated thinking skills necessary to perform competently in these disciplines.

By emphasizing problem solving and inquiry, many of the mathematics and science standards underscore the idea that students learn best when they are actively engaged in learning. Several of the standards also stress the need for students to build coherent structures of knowledge and be able to apply that knowledge in much the same manner as people who work in a particular discipline. For instance, the national science standards (NRC, 1996) state:

> Learning science is something students do, not something that is done to them. In learning science, students describe objects and events, ask questions, organize knowledge, construct explanations of natural phenomena, test those explanations in many different ways, and communicate their ideas to others. . . . Students establish connections between their current
knowledge of science and the scientific knowledge found in many sources; they apply science content to new questions; they engage in problem solving, planning, and group discussions; and they experience assessments that are consistent with an active approach to learning. (p. 20)

In these respects, the standards represent an important start toward incorporating findings from cognitive research about the nature of knowledge and expertise into curriculum and instruction. Standards vary widely, however, and some have fallen short of their intentions. For example, some state standards are too vague to be useful blueprints for instruction or assessment. Others call upon students to learn a broad range of content rather than focusing in depth on the most central concepts and methods of a particular discipline, and some standards are so detailed that the big ideas are lost or buried (American Federation of Teachers, 1999; Finn, Petrilli, and Vanourek, 1998).

State standards, whatever their quality, have significantly shaped classroom practices and exerted a major impact on assessment. Indeed, assessment is pivotal to standards-based reforms because it is the primary means of measuring progress toward attainment of the standards and of holding students, teachers, and administrators accountable for improvement over time. This accountability, in turn, is expected to create incentives for modifying and improving performance.

Without doubt, the standards movement has increased the amount of testing in K-12 schools and raised the consequences, expectations, and controversies attached to test results. To implement standards-based reforms, many states have put in place new tests in multiple curriculum areas and/or implemented tests at additional grade levels. Currently, 48 states have statewide testing programs, compared with 39 in 1996, and many school districts also have their own local testing programs (in addition to the range of classroom tests teachers regularly administer). As a result of this increased emphasis on assessment as an instrument of reform, the amount of spending on large-scale testing has doubled in the past 4 years, from $165 million in 1996 to $330 million in 2000 (Achieve, 2000).

Moreover, states and school districts have increasingly attached high stakes to test results. Scores on assessments are being used to make decisions about whether students advance to the next grade or graduate from high school, which students receive special services, how teachers and administrators are evaluated, how resources are allocated, and whether schools are eligible for various rewards or subject to sanctions or intervention by the district or state. These efforts have particular implications for equity if and when certain groups are disproportionately affected by the policies. As a result, the courts are paying greater attention to assessment results, and lawsuits are under way in several states that seek to use measures of educational quality to determine whether they are fulfilling their responsibility to provide all students with an adequate education (NRC, 1999c).
Although periodic testing is a critical part of any education reform, some of the movement toward increased testing may be fueled by a misguided assumption that more frequent testing, in and of itself, will improve education. At the same time, criticism of test policies may be predicated on an equally misguided assumption that testing, in and of itself, is responsible for most of the problems in education. A more realistic view is to address education problems not by stepping up the amount of testing or abandoning assessments entirely, but rather by refashioning assessments to meet current and future needs for quality information. However, it must be recognized that even very well-designed assessments cannot by themselves improve learning. Improvements in learning will depend on how well assessment, curriculum, and instruction are aligned and reinforce a common set of learning goals, and on whether instruction shifts in response to the information gained from assessments.

With so much depending on large-scale assessment results, it is more crucial than ever that the scores be reliable in a technical sense and that the inferences drawn from the results be valid and fair. It is just as important, however, that the assessments actually measure the kinds of competencies students need to develop to keep pace with the societal, economic, and technological changes discussed above, and that they promote the kinds of teaching and learning that effectively build those competencies. By these criteria, the heavy demands placed on many current assessments generally exceed their capabilities.

**Impact of Prior Theories of Learning and Measurement**

Current assessment practices are the cumulative product of theories of learning and models of measurement that were developed to fulfill the social and educational needs of a different time. This evolutionary process is described in more detail in Chapters 3 and 4. As Mislevy (1993, p. 19) has noted, “It is only a slight exaggeration to describe the test theory that dominates educational measurement today as the application of 20th century statistics to 19th century psychology.” Although the core concepts of prior theories and models are still useful for certain purposes, they need to be augmented or supplanted to deal with newer assessment needs.

Early standardized tests were developed at a time when enrollments in public schools were burgeoning, and administrators sought tools to help them educate the rapidly growing student populations more efficiently. As described in *Testing in American Schools* (U.S. Congress, Office of Technology Assessment, 1992), the first reported standardized written achievement exam was administered in Massachusetts in the mid-19th century and intended to serve two purposes: to enable external authorities to monitor school systems and to make it possible to classify children in pursuit of more efficient learning. Thus it was believed that the same tests used to monitor
the effectiveness of schools in accomplishing their missions could be used to sort students according to their general ability levels and provide schooling according to need. Yet significant problems have arisen in the history of assessment when it has been assumed that tests designed to evaluate the effectiveness of programs and schools can be used to make judgments about individual students. (Ways in which the purpose of an assessment should influence its design are discussed in Chapter 2 and more fully in Chapter 6.) At the same time, some educators also sought to use tests to equalize opportunity by opening up to individuals with high ability or achievement an educational system previously dominated by those with social connections—that is, to establish an educational meritocracy (Lemann, 1999). The achievement gaps that continue to persist suggest that the goal of equal educational opportunity has yet to be achieved.

Some aspects of current assessment systems are linked to earlier theories that assumed individuals have basically fixed dispositions to behave in certain ways across diverse situations. According to such a view, school achievement is perceived as a set of general proficiencies (e.g., mathematics ability) that remain relatively stable over situations and time.

Current assessments are also derived from early theories that characterize learning as a step-by-step accumulation of facts, procedures, definitions, and other discrete bits of knowledge and skill. Thus, the assessments tend to include items of factual and procedural knowledge that are relatively circumscribed in content and format and can be responded to in a short amount of time. These test items are typically treated as independent, discrete entities sampled from a larger universe of equally good questions. It is further assumed that these independent items can be accumulated or aggregated in various ways to produce overall scores.

**Limitations of Current Assessments**

The most common kinds of educational tests do a reasonable job with certain functions of testing, such as measuring knowledge of basic facts and procedures and producing overall estimates of proficiency for an area of the curriculum. But both their strengths and limitations are a product of their adherence to theories of learning and measurement that fail to capture the breadth and richness of knowledge and cognition. The limitations of these theories also compromise the usefulness of the assessments. The growing reliance on tests for making important decisions and for improving educational outcomes has called attention to some of their more serious limitations.

One set of concerns relates to whether the most widely used assessments effectively capture the kinds of complex knowledge and skills that are emphasized in contemporary standards and deemed essential for suc-
cess in the information-based economy described above (Resnick and Resnick, 1992; Rothman, Slattery, Vranek, and Resnick, in press). Traditional tests do not focus on many aspects of cognition that research indicates are important, and they are not structured to capture critical differences in students’ levels of understanding. For example, important aspects of learning not adequately tapped by current assessments include students’ organization of knowledge, problem representations, use of strategies, self-monitoring skills, and individual contributions to group problem solving (Glaser, Linn, and Bohrnstedt, 1997; NRC, 1999b).

The limits on the kinds of competencies currently being assessed also raise questions about the validity of the inferences one can draw from the results. If scores go up on a test that measures a relatively narrow range of knowledge and skills, does that mean student learning has improved, or has instruction simply adapted to a constrained set of outcomes? If there is explicit “teaching to the test,” at what cost do such gains in test scores accrue relative to acquiring other aspects of knowledge and skill that are valued in today’s society? This is a point of considerable current controversy (Klein, Hamilton, McCaffrey, and Stecher, 2000; Koretz and Barron, 1998; Linn, 2000).

A second issue concerns the usefulness of current assessments for improving teaching and learning—the ultimate goal of education reforms. On the whole, most current large-scale tests provide very limited information that teachers and educational administrators can use to identify why students do not perform well or to modify the conditions of instruction in ways likely to improve student achievement. The most widely used state and district assessments provide only general information about where a student stands relative to peers (for example, that the student scored at the 45th percentile) or whether the student has performed poorly or well in certain domains (for example, that the student performs “below basic in mathematics”). Such tests do not reveal whether students are using misguided strategies to solve problems or fail to understand key concepts within the subject matter being tested. They do not show whether a student is advancing toward competence or is stuck at a partial understanding of a topic that could seriously impede future learning. Indeed, it is entirely possible that a student could answer certain types of test questions correctly and still lack the most basic understanding of the situation being tested, as a teacher would quickly learn by asking the student to explain the answer (see Box 1-2). In short, many current assessments do not offer strong clues as to the types of educational interventions that would improve learners’ performance, or even provide information on precisely where the students’ strengths and weaknesses lie.

A third limitation relates to the static nature of many current assessments. Most assessments provide “snapshots” of achievement at particular points in time, but they do not capture the progression of students’ concep-
Consider the following two assessment situations:

**Assessment #1**
Question: What was the date of the battle of the Spanish Armada?  
Answer: 1588 [correct].  
Question: What can you tell me about what this meant?  
Answer: Not much. It was one of the dates I memorized for the exam. Want to hear the others?

**Assessment #2**
Question: What was the date of the battle of the Spanish Armada?  
Answer: It must have been around 1590.  
Question: Why do you say that?  
Answer: I know the English began to settle in Virginia just after 1600, not sure of the exact date. They wouldn’t have dared start overseas explorations if Spain still had control of the seas. It would take a little while to get expeditions organized, so England must have gained naval supremacy somewhere in the late 1500s.

Most people would agree that the second student showed a better understanding of the Age of Colonization than the first, but too many examinations would assign the first student a better score. When assessing knowledge, one needs to understand how the student connects pieces of knowledge to one another. Once this is known, the teacher may want to improve the connections, showing the student how to expand his or her knowledge.

---

**BOX 1.2 Rethinking the Best Ways to Assess Competence**

Consider the following two assessment situations:

**Assessment #1**
Question: What was the date of the battle of the Spanish Armada?  
Answer: 1588 [correct].  
Question: What can you tell me about what this meant?  
Answer: Not much. It was one of the dates I memorized for the exam. Want to hear the others?

**Assessment #2**
Question: What was the date of the battle of the Spanish Armada?  
Answer: It must have been around 1590.  
Question: Why do you say that?  
Answer: I know the English began to settle in Virginia just after 1600, not sure of the exact date. They wouldn’t have dared start overseas explorations if Spain still had control of the seas. It would take a little while to get expeditions organized, so England must have gained naval supremacy somewhere in the late 1500s.

Most people would agree that the second student showed a better understanding of the Age of Colonization than the first, but too many examinations would assign the first student a better score. When assessing knowledge, one needs to understand how the student connects pieces of knowledge to one another. Once this is known, the teacher may want to improve the connections, showing the student how to expand his or her knowledge.
reviews to monitor for this kind of bias. The use of assessments for high-stakes decisions raises additional questions about fairness (NRC, 1999c). If the assessments are not aligned with what students are being taught, it is not fair to base promotion or rewards on the results, especially if less advantaged students are harmed disproportionately by the outcome. If current assessments do not effectively measure the impact of instruction or fail to capture important skills and knowledge, how can educators interpret and address gaps in student achievement?

One of the main goals of current reforms is to improve learning for low-achieving students. If this goal is to be accomplished, assessment must give students, teachers, and other stakeholders information they can use to improve learning and inform instructional decisions for individuals and groups, especially those not performing at high levels. To be sure, assessments by themselves do not cause or cure inequities in education; indeed, many of the causes of such inequities are beyond the scope of the education system itself. However, when assessment fails to provide information that can enhance learning, it leaves educators ill equipped to close achievement gaps.

While concerns associated with large-scale tests have received considerable attention, particularly in recent years, the classroom assessments commonly used by teachers also are often limited in the information they provide. Just as large-scale tests have relied on an incomplete set of ideas about learning, so, too, have the kinds of assessments teachers regularly administer in their classrooms. Often, teachers adhere to assessment formats and scoring practices found in large-scale tests. This can be traced largely to teacher education programs and professional development experiences that have for the most part failed to equip teachers with contemporary knowledge about learning and assessment, especially the knowledge needed to develop tasks that would elicit students’ thinking skills or make it possible to assess their growth and progress toward competence (Gizek, 2000; Dwyer, 1998).

Alternative Assessment Practices

Standards-based reform continues to stimulate research and development on assessment as people seek to design better approaches for measuring valued knowledge and skills. States and school districts have made major investments to better align tests with standards and to develop alternative approaches for assessing knowledge and skills not well captured by most current tests. Teachers have been offered professional development opportunities focusing on the development and scoring of new state assessment instruments more closely aligned with curricular and instructional practices. Nowhere has this confluence of activity been more evident than in the area of “performance assessment” (Council of Chief State School Officers, 1999;

The quest for alternatives to traditional assessment modes has led many states to pursue approaches that include the use of more open-ended tasks that call upon students to apply their knowledge and skills to create a product or solve a problem. Performance assessment represents one such effort to address some of the limitations of traditional assessments. Performance assessment, an enduring concept (e.g., Lindquist, 1951) that attracted renewed attention during the 1990s, requires students to perform more “authentic” tasks that involve the application of combined knowledge and skills in the context of an actual project. Even with such alternative formats, however, there has been a constant gravitation toward familiar methods of interpreting student performance. For example, Baxter and Glaser (1998) analyzed a range of current performance assessments in science and often found mismatches between the intentions of the developers and what the tasks and associated scoring rubrics actually measured. Particularly distressing was the observation that some performance tasks did not engage students in the complex thinking processes intended.

As a result of these limitations, the growing interest in performance assessment was followed by a recognition that it is not the hoped-for panacea, especially in light of the costs, feasibility, and psychometric concerns associated with the use of such measures (Mehrens, 1998; National Center for Education Statistics, 1996). The cumulative work on performance assessment serves as a reminder that the key question is whether an assessment, whatever its format, is founded on a solid model of learning and whether it will provide teachers and students with information about what students know that can be used for meaningful instructional guidance.

Simply put, steps have been taken to improve assessment, but a significant leap forward needs to occur to equip students, parents, teachers, and policy makers with information that can help them make appropriate decisions about teaching practices and educational policies that will assist learning. Fortunately, the elements of change that could produce such an advance are already present within the cognitive and measurement sciences.

**Assessment Based on Contemporary Foundations**

Several decades of research in the cognitive sciences has advanced the knowledge base about how children develop understanding, how people reason and build structures of knowledge, which thinking processes are associated with competent performance, and how knowledge is shaped by social context. These findings, presented in Chapter 3, suggest directions for revamping assessment to provide better information about students’ levels
of understanding, their thinking strategies, and the nature of their misunderstandings.

During this same period, there have been significant developments in measurement methods and theory. As presented in Chapter 4, a wide array of statistical measurement methods is currently available to support the kinds of inferences that cognitive research suggests are important to draw when measuring student achievement.

In this report we describe examples of some initial and promising attempts to capitalize on these advances. However, these efforts have been limited in scale and have not yet coalesced around a set of guiding principles. In addition to discerning those principles, it is necessary to undertake more research and development to move the most promising ideas and prototypes into the varied and unpredictable learning environments found in diverse classrooms embedded within complex educational systems and policy structures.

In pursuing new forms of assessment, it is important to remember that assessment is a system composed of the three interconnected elements discussed earlier—cognition, observation, and interpretation—and that assessments function within a larger system of curriculum, instruction, and assessment. Radically changing one of these elements and not the others runs the risk of producing an incoherent system. All of the elements and how they interrelate must be considered together.

Moreover, while new forms of assessment could address some of the limitations described above and give teachers, administrators, and policy makers tools to help them improve schooling, it is important to note that tests by themselves do not improve teaching and learning, regardless of how effective they are at providing information about student competencies. Many factors affect instruction and learning, including the quality of the curriculum, the experience and skills of teachers, and the support students receive outside of class. It is also essential to keep in mind that any assessment operates within constraints, and these constraints can limit its ability to provide useful information. For example, such factors as the amount of money available for developing an assessment and the amount of instructional time available for its administration or scoring can restrict the types of tasks used for the assessment and thus the evidence it can provide about student learning. In addition, classroom factors such as class size and opportunity for teachers to interact with one another can affect teachers’ ability to profit from the information that is derived. Thus while new assessments can enhance the available information about student competencies, their full potential can be realized only by removing such constraints.

That potential is significant. Assessments that inform teachers about the nature of student learning can help them provide better feedback to students, which in turn can enhance learning (Black and Wiliam, 1998). Assess-
ments based on theories of how competence develops across grade levels in a curriculum domain could provide more valid measures of growth and the value added by teachers and schools.

Assessments based on current cognitive principles and measurement theories could also enhance community dialogue about goals for student learning and indicators of achievement at various grade levels and in different subject areas. Comparisons based on attainment of worthwhile learning goals, rather than normative descriptions of how students perform, could enhance the public’s understanding of educational quality. New forms of assessment could also help provide descriptive and accurate information about the nature of achievement in a subject area and patterns of students’ strengths and weaknesses that would be more useful than existing data for guiding policy decisions and reform efforts.

Issues of fairness and equity must be central concerns in any effort to develop new forms of assessment. Relevant to these issues is a substantial body of research on the social and cultural dimensions of cognition and learning (discussed in Chapter 3). To improve the fairness of assessment, it must be recognized that cultural practices equip students differently to participate in the discourse structures that are often unique to testing contexts. It is all too easy to conclude that some cultural groups are deficient in academic competence, when the differences can instead be attributable to cultural variations in the ways students interpret the meaning, information demands, and activity of taking tests (Steele, 1995, 1997). These sorts of differences need to be studied and taken into account when designing and interpreting the results of assessments. If well-designed and used, new models of assessment could not only measure student achievement more fairly, but also promote more equitable opportunity to learn by providing better-quality information about the impact of educational interventions on children. More informative classroom assessments could result in earlier identification of learning problems and interventions for children at risk, rather than waiting for results from large-scale assessments to signal problems. Students with disabilities could also benefit from this approach. At the same time, it will be necessary for educators and researchers to monitor the effects of their practices continually to ensure that new assessments do not exacerbate existing inequalities. While there are many reasons to question the fairness of current testing practices, it would be misguided to implement new assessment approaches and assume that they promote fairness and equity without validating such presumptions.

ISSUES AND CHALLENGES

The key issues that emerge from the themes discussed above strongly suggest that it is appropriate and necessary to rethink the scientific prin-
ciples and philosophical assumptions that serve as the foundations of educational assessment. Doing so will provide new ways of understanding and approaching these issues and finding solutions to the assessment challenges they pose.

- Expectations about what all students should learn—and, by implication, what they should be tested on—have changed in response to social, economic, and technological changes and as a result of the standards-based reform movement. All students are now expected to demonstrate the kinds of reasoning and problem-solving abilities once expected of only a minority of young people. Assessments are needed to gauge these aspects of student competence.
- Standards-based reform has increased both the amount of testing and the stakes attached to test results. This development has placed more pressure on current assessment systems than they were meant to bear and has highlighted some of their limitations.
- Current assessment systems are the cumulative product of various prior theories of learning and methods of measurement. Although some of these foundations are still useful for certain functions of testing, change is needed. Assessment systems need to evolve to keep pace with developments in the sciences of learning and measurement and to achieve the learning goals pursued by reformers.
- Four decades of research in the cognitive sciences has advanced the knowledge base about how children develop understanding, how people reason and build structures of knowledge, which thinking processes are associated with competent performance, and how knowledge is shaped by social context. These findings suggest directions for revamping assessment to enable more valid and fair inferences about students’ levels of understanding, their thinking strategies, and the nature of their misunderstandings.
- Developments in the science and technology of assessment have made available a variety of measurement methods and statistical models that could be used to design assessments capable of better capturing the complexity of cognition and learning.
- A science of assessment that brought together cognitive principles and highly developed measurement models could address some of the limitations of current assessments and yield a number of benefits for students, teachers, and the educational system as a whole. Effort must be made to study what can be accomplished through programs of sustained assessment design and implementation based on current scientific knowledge.
- At the same time, it is important to recognize that any assessment operates within the constraints of the larger education system. The ability of new forms of assessment to function to their fullest potential can be im-
peded by constraints such as limited resources and time for assessment; large class sizes and little time for teachers to interact; and misalignment among curriculum, instruction, and assessment. The influences of such factors can not be ignored but must be incorporated into the process of assessment reform.

Education reform will be difficult to achieve if educators continue to carry the weight of practices designed for times past. New methods of assessment can begin to drive changes in curriculum, teaching, and learning that support patterns of human cognitive growth and prepare people for dignified lives, workplace competence, and social development.

**STRUCTURE OF THIS REPORT**

This report addresses many of the conceptual issues and pragmatic challenges noted above. It is divided into four parts, as detailed below. Part I consists of this chapter and Chapter 2, which provides background on the purposes and nature of assessment and introduces key concepts used throughout the report.

Part II, consisting of Chapters 3 and 4, explains how expanding knowledge in the fields of human cognition and measurement can form the foundations for an improved approach to assessment. Chapter 3 reviews contemporary understanding of how people learn, focusing on findings that have implications for improving educational assessment. The discussion addresses the way knowledge is represented and organized in the mind, the characteristics of expertise in a discipline and the development of that expertise, and the influence of cultural and social factors on learning. Chapter 4 describes current measurement methods, both familiar and new, and why they evolved. It explores how the broad array of existing methods can be used to develop a new generation of assessments that can provide better evidence of students’ understanding and cognitive processes.

Part III, consisting of Chapters 5, 6, and 7, sets forth principles for designing and using assessments based on advances in cognitive and measurement theories. Chapter 5 describes features of a new approach to assessment design based on a synthesis of cognitive and measurement principles; existing and innovative assessment examples are used to illustrate the application of the general design principles to different assessment purposes and contexts. The discussion focuses on how current educational testing guidelines and practice could be improved by making stronger connections between advances in cognitive and measurement theories. Chapter 6 addresses contrasts and design trade-offs between classroom and large-scale assessment, and explores how assessments can be designed and used in each context to improve student learning. Opportunities for enhancing the synergy between classroom and large-scale assessment are also addressed. Chap-
ter 7 considers the role of technology in transforming both the kinds of learning that should be assessed and the assessment methods used. The chapter includes examples of technological tools that illustrate new uses for assessment and highlights some issues that need to be considered as technology becomes more important in education.

Chapters 2 through 7 open with a listing of the themes used to organize the discussion that follows. Each of these chapters ends with a set of conclusions based on the findings and analysis presented under those themes.

Part IV, Chapter 8, proposes a research and development agenda for expanding the knowledge base on the integration of cognition and measurement. It also considers the avenues through which the growing knowledge base is most likely to have an impact on actual assessment practice.