Part IV

Conclusion
Implications and Recommendations for Research, Policy, and Practice

The Committee on the Foundations of Assessment produced this report, with the support of the National Science Foundation (NSF), to review and synthesize advances in the cognitive and measurement sciences and to explore the implications of those advances for improving educational assessment. Interest in the intersection of these two fields is not new. Prompted by calls for assessments that can better inform and support learning, a number of education researchers have put forth the potential benefits of merging modern knowledge in the areas of cognition and learning and methods of educational measurement (Baker, 1997; Cronbach and Gleser, 1965; Glaser, 1981; Glaser and Silver, 1994; Messick, 1984; Mislevy, 1994; National Academy of Education, 1996; Nichols, 1994; Pellegrino, Baxter, and Glaser, 1999; Snow and Lohman, 1993; Wilson and Adams, 1996).

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 (National Research Council [NRC], 1999b). 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 (psychometric) methods and theory. As presented in Chapter 4, a wide array of statistical measurement methods are currently available to support the kinds of inferences that cognitive research suggests are important to pursue when assessing student achievement.

Meanwhile, computer and telecommunications technologies are making it possible to assess what students are learning at very fine levels of detail,
with vivid simulations of real-world situations, and in ways that are tightly integrated with instruction. Chapter 7 provides examples of how technology is making it feasible, for instance, for students to receive ongoing individualized feedback as they work with a computerized tutoring system—feedback more detailed than what a teacher could have provided a class of 30 students in the past.

This report describes a variety of promising assessment innovations that represent first steps in capitalizing on these opportunities. However, most of these examples have been limited to small-scale applications that have yet to affect mainstream assessment practice. In this final chapter, we discuss priorities for research, practice, and policy to enable the emergence of a "new science of assessment." First, however, we summarize some of the main points from the preceding chapters by describing a vision for a future generation of educational assessments based on the merger of modern cognitive theory and methods of measurement.

**A VISION FOR THE FUTURE OF ASSESSMENT**

In the future envisioned by the committee, educational assessments will be viewed as a facilitator of high levels of student achievement. They will help students learn and succeed in school by making as clear as possible to them, their teachers, and other education stakeholders the nature of their accomplishments and the progress of their learning.

Teachers will assess students’ understanding frequently in the classroom to provide them with feedback and determine next steps for instruction. Their classroom practices will be grounded in principles of how students think and learn in content domains and of assessment as a process of reasoning from evidence. Teachers will use this knowledge to design assessments that provide students with feedback about particular qualities of their work and what they can do to improve.

Students will provide evidence of their understanding and thinking in a variety of ways—by responding to teachers’ questions, writing or producing projects, working with computerized tutoring systems, or attempting to explain concepts to other students. Teachers, in turn, will use this information to modify instruction for the class and for individuals on the basis of their understanding and thinking patterns.

Teachers will have a clear picture of the learning goals in subject domains, as well as typical learning pathways for reaching those goals. Ultimate and intermediate learning goals will be shared regularly with students as a part of instruction. Students will be engaged in activities such as peer and self-assessment to help them internalize the criteria for high-quality work and develop metacognitive skills.
Teachers will also use summative assessments for ongoing reflection and feedback about overall progress and for reporting of this information to others. External summative assessments, such as state tests, will reinforce the same ultimate goals and beliefs about learning that are operating in the classroom. Large-scale assessments will set valuable learning goals for students to pursue. Such assessments will broadly sample the desired outcomes for learning by using a variety of methods, such as on-demand assessment combined with a sampling of work produced during the course of instruction.

Policy makers, educators, and the public will come to expect more than the general comparisons and rankings that characterize current test results. Performance on large-scale assessments will be explicitly and publicly displayed so that students, parents, and teachers can see the concepts and processes entailed at different levels of competence. Assessments will be able to show, for instance, how a competent performer proceeds on a mathematics problem and forms an answer, in comparison with a student who is less proficient. Large-scale assessments will help show the different kinds of interpretations, procedural strategies, explanations, and products that differentiate among various levels or degrees of competence.

Within an education system, teachers, administrators, and policy makers will be working from a shared knowledge base about how students learn subject matter and what aspects of competence are important to assess. Resource materials that synthesize modern scientific understanding of how people learn in areas of the curriculum will serve as the basis for the design of classroom and large-scale assessments, as well as curriculum and instruction, so that all the system’s components work toward a coherent set of learning goals.

In many ways, this vision for assessment represents a significant departure from the types of assessments typically available today and from the ways in which such assessments are most commonly used. Current knowledge could serve as the basis for a number of improvements to the assessment design process (as described in Chapters 3, 4, and 5 of this report) to produce assessment information that would be more useful, valid, and fair. Full realization of the committee’s broader vision for educational assessment, however, will require more knowledge about how to design and use such assessments, as well as about the underlying fundamental properties of learning and measurement. Furthermore, the committee recognizes that the maximum potential of new forms of assessment cannot be realized unless educational practices and policies adapt in significant ways. Some of the constraints that currently limit assessment practice will need to be relaxed if the full benefits of a merger between the cognitive and measurement sciences are to be realized. The new kinds of assessment described in this report do not necessarily conform to the current mode of on-demand, pa-
per-and-pencil tests that students take individually at their desks under strictly standardized conditions. Furthermore, realizing the potential benefits of new forms of assessment will depend on making compatible changes in curriculum and instruction.

**BRIDGING RESEARCH AND PRACTICE**

Like other groups before us (NRC, 1999c; National Academy of Education, 1999), the committee recognizes that the bridge between research and practice takes time to build and that research and practice must proceed interactively. It is unlikely that the insights gained from current or new knowledge about cognition, learning, and measurement will be sufficient by themselves to bring about transformations in assessment such as those described in this report. As the NRC’s Committee on Learning Research and Educational Practice pointed out, research and practice need to be connected more directly through the building of a cumulative knowledge base that serves both sets of interests. In the context of this study, that knowledge base would focus on the development and use of theory-based assessment. Furthermore, it is essential to recognize that research impacts practice indirectly through the influence of the existing knowledge base on four important mediating arenas: educational tools and materials; teacher education and professional development; education policies; and public opinion and media coverage (NRC, 1999c). By affecting each of these arenas, an expanding knowledge base on the principles and practices of effective assessment can help change educational practice. And the study of changes in practice, in turn, can help in further developing the knowledge base. These organizing ideas regarding the connections between research and practice are illustrated in Figure 8-1.

In this chapter we outline a proposed research and development agenda for expanding the knowledge base on the integration of cognition and measurement and consider the implications of such a knowledge base for each of the four mediating arenas that directly influence educational practice. In doing so we propose two general guidelines for how future work should proceed.

*First, the committee advocates increased and sustained multidisciplinary collaboration around theoretical and practical matters of assessment.* We apply this precept not only to the collaboration between researchers in the cognitive and measurement sciences, but also to the collaboration of these groups with teachers, curriculum specialists, and assessment developers. The committee believes the potential for an improved science and design of educational assessment lies in a mutually catalytic merger of the two foundational disciplines, especially as such knowledge is brought to bear on conceptual and pragmatic problems of assessment development and use.
Second, the committee urges individuals in multiple communities, from research through practice and policy, to consider the conceptual scheme and language used in this report as a guide for stimulating further thinking and discussion about the many issues associated with the productive use of assessments in education. The assessment triangle set forth in Chapter 2 and summarized in Box 8-1 provides a conceptual framework for principled thinking about the assumptions and foundations underlying an assessment. In the next section of this chapter we consider some of the implications of our conceptual scheme for research that can contribute to the advancement of both theory and practice.

Before discussing specific implications for research and practice and presenting our recommendations in each of these areas, we would be remiss if we did not note our concern about continuing with the present system of educational assessment, including the pattern of increasing investment in large-scale assessment designs and practices that have serious limi-
The process of reasoning from evidence can be portrayed as a triangle referred to throughout this report as the assessment triangle. As shown below, the corners of the triangle represent three key elements that underlie any assessment: (1) a model of student cognition and learning in the domain, (2) a set of beliefs about the kinds of observations that will provide evidence of students’ competencies, and (3) an interpretation process for making sense of the evidence.

These three elements form the foundations on which every assessment rests. The three elements are represented as corners of a triangle because each is connected to and dependent on the others. To have an effective assessment, all three should be explicitly coordinated as part of the design. A major tenet of this report is that most assessments in current use are based on outmoded conceptions of cognition and learning and on impoverished observation and interpretation methods, as compared with what could be the case given modern scientific knowledge of cognition and measurement.
IMPLICATIONS AND RECOMMENDATIONS FOR RESEARCH, POLICY, AND PRACTICE

8 IMPLICATIONS AND RECOMMENDATIONS FOR RESEARCH, POLICY, AND PRACTICE

IMPLICATIONS AND RECOMMENDATIONS FOR RESEARCH

The research needed to approach the new science of assessment envisioned by the committee needs to focus on those issues that lie at the intersection of cognitive and measurement science. In this section we present the committee’s recommendations for research organized into three broad categories: (1) synthesis of existing knowledge, (2) research to expand the current knowledge base, and (3) some initial steps for building the knowledge base.

For all the research recommendations presented below, we advocate a general approach to research and development that differs from conventional practices. In the traditional view of research, development, and implementation, scientists begin with basic research that involves gathering fundamental knowledge and developing theories about an area of inquiry. Other scientists and practitioners use this basic research, together with their experience, to design prototypes that apply the knowledge in practical settings. Still others then design ways to implement the prototypes on a larger scale.

The committee believes that, in the case of the assessments we envision, research should focus on design and implementation. The committee takes this position for two reasons. The first is strategic. As described throughout this report, some promising prototype assessments based on modern cognitive theory and measurement principles have already been developed. While the prototypes have been used effectively in selected classrooms and educational settings, there is generally limited experience with their application outside of relatively controlled settings or in large-scale contexts. In part this is because the new forms of assessment are often complex and have not been tailored for widespread practical use. In addition, there are issues involved in large-scale assessment that designers of classroom-based tools...
have yet to confront. The committee takes the position that practical implementation should be studied to raise questions about fundamental science.

In his book *Pasteur’s Quadrant*, Stokes (1997) argues that the traditional dichotomy between “basic” and “applied” research is not always applicable. In many instances, research aimed at solving practical problems can test the validity and generality of fundamental principles and knowledge. Pasteur’s work is an archetype of this approach. By focusing on a very real practical problem—developing ways to combat harmful bacteria—Pasteur pursued “use-inspired strategic research” that not only helped solve the immediate problem, but also contributed greatly to enhancing fundamental knowledge about biology and biochemistry. Similarly, Hargreaves (1999) argues that research results cannot be applied directly to classroom practice, but must be transformed by practitioners; that is, teachers need to participate in creating new knowledge.

In a report to the National Education Research Policies and Priorities Board of the Office of Educational Research and Improvement, a panel of the National Academy of Education argues that federal agencies should fund research in Pasteur’s Quadrant as well as basic research (National Academy of Education, 1999). The panel states that “problem-solving research and development” (the equivalent of what Stokes describes as use-inspired strategic research) is characterized by four features:

- Commitment to the improvement of complex systems.
- Co-development by researchers and practitioners, with recognition of differences in expertise and authority.
- Long-term engagement that involves continual refinement.
- Commitment to theory and explanation.

The panel notes that this last feature would enable prototypes generated in one site or context of use to “travel” to other settings (the panel contrasts its view with the traditional notion of “dissemination”). To permit wider adoption, the research would have to generate principles to ensure that others would not simply replicate the surface features of an innovation. Also required would be consideration of tools that could help others apply the innovation faithfully, as well as people familiar with the design who could help others implement it. The committee is sympathetic to this argument and believes research that addresses ways to design assessments for use in either classrooms or large-scale settings can simultaneously enhance understanding of the design principles inherent in such assessments and improve basic knowledge about cognition and measurement.

We advocate that the research recommended below be funded by federal agencies and private foundations that currently support research on teaching and learning, as well as private-sector entities involved in com-
cial assessment design and development. Among the salient federal agencies are the Department of Education, the NSF, and the National Institute of Child Health and Human Development. The research agenda is expansive in both scope and likely duration. It would be sensible for the funding of such work to be coordinated across agencies and, in many instances, pursued cooperatively with foundations and the private sector.

### Synthesis of Existing Knowledge

**Recommendation 1:** Accumulated knowledge and ongoing advances from the merger of the cognitive and measurement sciences should be synthesized and made available in usable forms to multiple educational constituencies. These constituencies include educational researchers, test developers, curriculum specialists, teachers, and policy makers.

As discussed throughout this report, a great deal of the foundational research needed to move the science of assessment forward has already been conducted; however, it is not widely available or usable in synthetic form. This report is an initial attempt at such a synthesis, but the committee recognized from the start of its work that a comprehensive critique, synthesis, and extrapolation of all that is known was beyond the scope of a study such as this and remains a target for the future. Furthermore, there is an ongoing need to accumulate, synthesize, and disseminate existing knowledge—that is, to construct the cumulative knowledge base on assessment design and use that lies at the center of Figure 8-1.

### Expanding the Knowledge Base

**Recommendation 2:** Funding should be provided for a major program of research, guided by a synthesis of cognitive and measurement principles, focused on the design of assessments that yield more valid and fair inferences about student achievement. This research should be conducted collaboratively by multidisciplinary teams comprising both researchers and practitioners.

- A priority should be the development of cognitive models of learning that can serve as the basis for assessment design for all areas of the school curriculum. Research on how students learn subject matter should be conducted in actual educational settings and with groups of learners representative of the diversity of the student population to be assessed.
• Research on new statistical measurement models and their applicability should be tied to modern theories of cognition and learning. Work should be undertaken to better understand the fit between various types of cognitive theories and measurement models to determine which combinations work best together.

• Research on assessment design should include exploration of systematic and fair methods for taking into account aspects of examinees’ instructional background when interpreting their responses to assessment tasks. This research should encompass careful examination of the possible consequences of such adaptations in high-stakes assessment contexts.

One priority for research is the development of cognitive models of learning for areas of the school curriculum. As noted in Chapter 3, researchers have developed sophisticated models of student cognition in various areas of the curriculum, such as algebra and physics. However, an understanding of how people learn remains limited for many other areas. Moreover, even in subject domains for which characteristics of expertise have been identified, a detailed understanding of patterns of growth that would enable one to identify landmarks on the way to competence is often lacking. Such landmarks are essential for effective assessment design and implementation.

The development of models of learning should not be done exclusively by scientists in laboratory settings. As argued earlier, it would be more fruitful if such investigations were conducted, at least in part, in actual educational contexts by collaborative teams of researchers and practitioners. Such collaborations would help enhance both the quality and utility of the knowledge produced by the research.

To develop assessments that are fair—that are comparably valid across different groups of students—it is crucial that patterns of learning for different populations of students are studied. Much of the development of cognitive theories has been conducted with a restricted group of students (i.e., mostly middle-class whites). In many cases it is not clear whether current theories of learning apply equally well with diverse populations of students, including those who have been poorly served in the educational system, underrepresented minority students, English-language learners, and students with disabilities. There are typical learning pathways, but not a single pathway to competence. Furthermore, students will not necessarily respond in similar ways to assessment probes designed to diagnose knowledge and understanding. These kinds of natural variations among individuals need to
be better understood through empirical study and incorporated into the cognitive models of learning that serve as a basis for assessment design.

Sophisticated models of learning by themselves do not produce high-quality assessment information. Also needed are methods and tools both for eliciting appropriate and relevant data from students and for interpreting the data collected about student performance. As described in Chapter 4, the measurement methods now available enable a much broader range of inferences to be drawn about student competence than many people realize. But research is needed to investigate the relative utility of existing and future statistical models for capturing critical aspects of learning specified in cognitive theories.

Most of the new measurement models have been applied only on a limited scale. Thus, there is a need to explore the utility and feasibility of the new models for a wider range of assessment applications and contexts. Within such a line of inquiry, a number of issues will need to be understood in more depth, including the level of detail at which models of student learning must be specified for implementing various types of classroom or large-scale assessments. Furthermore, there is a vital need for research on ways to make a broader range of measurement models usable by practitioners, rather than exclusively by measurement specialists. Many of the currently available measurement methods require complex statistical modeling that only people with highly specialized technical skills can use to advantage. If these tools are to be applied more widely, understandable interfaces will need to be built that rise above statistical complexity to enable widespread use, just as users of accounting and management programs need not understand all the calculations that go into each element of the software.

Another priority for assessment design is the exploration of new ways to address persisting issues of fairness and equity in testing. People often view fairness in testing in terms of ensuring that students are placed in test situations that are as similar or standardized as possible. But another way of approaching fairness is to take into account examinees' histories of instruction or opportunities to learn the material being tested when interpreting their responses to assessment tasks. Ways of drawing such conditional inferences have been tried mainly on a small scale but hold promise for tackling persisting issues of equity in assessment.

**Recommendation 3:** Research should be conducted to explore how new forms of assessment can be made practical for use in classroom and large-scale contexts and how various new forms of assessment affect student learning, teacher practice, and educational decision making.

- Research should explore ways in which teachers can be assisted in integrating new forms of assessment into their in-
structional practices. It is particularly important that such work be done in close collaboration with practicing teachers who have varying backgrounds and levels of teaching experience.

• Also to be studied are ways in which school structures (e.g., length of time of classes, class size, and opportunity for teachers to work together) impact the feasibility of implementing new types of assessments and their effectiveness.

The committee firmly believes that the kinds of examples described in this report—all of which are currently being used in classrooms or large-scale contexts—represent positive steps toward the development of assessments that can not only inform but also improve learning. However, for these kinds of innovations to gain more widespread adoption, work is needed to make them practical for use in classroom and large-scale contexts, and evidence of their impact on student learning is needed.

Furthermore, the power offered by assessments to enhance learning in large numbers of classrooms depends on changes in the relationship between teacher and student, the types of lessons teachers use, the pace and structure of instruction, and many other factors. To take advantage of the new tools, many teachers will have to change their conception of their role in the classroom. They will have to shift toward placing much greater emphasis on exploring students’ understanding with the new tools and then undertaking a well-informed application of what has been revealed by use of the tools. This means teachers must be prepared to use feedback from classroom and external assessments to guide their students’ learning more effectively by modifying the classroom and its activities. In the process, teachers must guide their students to be more engaged actively in monitoring and managing their own learning—to assume the role of student as self-directed learner.

The power of new assessments depends on substantial changes not only in classroom practice, but also in the broader educational context in which assessments are conducted. For assessment to serve the goals of learning, there must be alignment among curriculum, instruction, and assessment. Furthermore, the existing structure and organization of schools may not easily accommodate the type of instruction users of the new assessments will need to employ. For instance, if teachers are going to gather more assessment information during the course of instruction, they will need time to assimilate that information. If these kinds of systemic and structural issues are not addressed, new forms of assessment will not live up to their full potential. This is a common fate for educational innovations. Many new techniques and procedures have failed to affect teaching and learning on a large scale because the innovators did not address all the factors that affect
teaching and learning (Elmore, 1996). Despite the promise of new procedures, most teachers tend to teach the way they have always taught, except in the “hothouse” settings where the innovations were designed.

Thus, if assessments based on the foundations of cognitive and measurement science are to be implemented on a broad scale, changes in school structures and practices will likely be needed. However, the precise nature of such changes is uncertain. As new assessments are implemented, researchers will need to examine the effects of such factors as class size and the length of the school day on the power of assessments to inform teachers and administrators about student learning. Also needed is a greater understanding of what structural changes are required for teachers to modify their practice in ways that will enable them to incorporate such assessments effectively.

Some Initial Steps for Building the Knowledge Base

Recommendation 4: Funding should be provided for in-depth analyses of the critical elements (cognition, observation, and interpretation) underlying the design of existing assessments that have attempted to integrate cognitive and measurement principles (including the multiple examples presented in this report). This work should also focus on better understanding the impact of such exemplars on student learning, teacher practice, and educational decision making.

The committee believes an ideal starting point for much of the research agenda is further study of the types of assessment examples provided in the preceding chapters, which represent initial attempts at synthesizing advances in the cognitive and measurement sciences. While these examples were presented to illustrate features of the committee’s proposed approach to assessment, the scope of this study did not permit in-depth analyses of all the design and operational features of each example or their impact on student learning, teacher practice, and educational decision making. Further analysis of these and other examples would help illuminate the principles and practices of assessment design and use described in this report. Several important and related directions of work need to be pursued.

First, to fully understand any assessment, one must carefully deconstruct and analyze it in terms of its underlying foundational assumptions. The assessment triangle provides a useful framework for analyzing the foundational elements of an assessment. Questions need to be asked and answered regarding the precise nature of the assumptions made about cognition, observation, and interpretation, including the degree to which they are in synchrony. Such an analysis should also consider ways in which current knowl-
edge from the cognitive and measurement sciences could be used to enhance the assessment in significant ways.

Second, once an assessment is well understood, its effectiveness as a tool for measurement and for support of learning must be explored and documented. The committee strongly believes that the examples in this report represent promising directions for further development, and where available, has presented empirical support for their effectiveness. However, there is a strong need for additional empirical studies aimed at exploring which tools are most effective and why, how they can best be used, and what costs and benefits they entail relative to current forms of assessment.

Third, while it is important to carefully analyze each of the examples as a separate instance of innovative design, they also need to be analyzed as a collective set of instances within a complex “design space.” The latter can be thought of as a multivariate environment expressing the important features that make specific instances simultaneously similar and different. This design space is only partially conceived and understood at the present time. Thus, analyses should be pursued that cut across effective exemplars with the goal of identifying and clarifying the underlying principles of the new science of assessment design. In this way, the principles described in this report can be refined and elaborated while additional principles and operational constructs are uncovered. If a new science of assessment grounded in concepts from cognitive and measurement science is to develop and mature, every attempt must be made to uncover the unique elements that emerge from the synthesis of the foundational sciences. This work can be stimulated by further in-depth analysis of promising design artifacts and the design space in which they exist.

Recommendation 5: Federal agencies and private-sector organizations concerned about issues of assessment should support the establishment of multidisciplinary discourse communities to facilitate cross-fertilization of ideas among researchers and assessment developers working at the intersection of cognitive theory and educational measurement.

Many of the innovative assessment practices described in this report were derived from projects funded by the NSF or the James S. McDonnell Foundation. These organizations have provided valuable opportunities for cross-fertilization of ideas, but more sharing of knowledge is needed. Many of the examples exist in relative isolation and are known only within limited circles of scientific research and/or educational practice. The committee believes there are enough good examples of assessments based on a merger of the cognitive and measurement sciences so that designers can start building from existing work. However, a discourse among multidisciplinary commu-
nities will need to be established to promote and sustain such efforts. As mentioned earlier, this report provides a language and conceptual base for discussing the ideas embedded in existing innovative assessment practices and for the broader sharing and critique of those ideas.

**IMPLICATIONS AND RECOMMENDATIONS FOR POLICY AND PRACTICE**

Research often does not directly affect educational practice, but it can effect educational change by influencing the four mediating arenas of the education system that do influence practice, shown previously in Figure 8-1. For the earlier committee that identified these arenas, the question was how to bridge research on student learning and instructional practice in classrooms. The focus of the present committee is on a related part of the larger question: how to link research on the integration of cognition and measurement with actual assessment practice in schools and classrooms. By influencing and working through the four mediating arenas, the growing knowledge base on cognition and measurement can ultimately have an effect on assessment and instructional practice in classrooms and schools.

It is important to note that the path of influence does not flow only in one direction. Just as we believe that research on the integration of cognition and measurement should focus on use-inspired strategic research, we believe that practical matters involving educational tools and materials, teacher education and professional development, education policies, and public opinion and media coverage will influence the formulation of research questions that can further contribute to the development of a cumulative knowledge base. Research focused on these arenas will enhance understanding of practical matters related to how students learn and how learning can best be measured in a variety of school subjects.

**Educational Tools and Materials**

Recommendation 6: Developers of assessment instruments for classroom or large-scale use should pay explicit attention to all three elements of the assessment triangle (cognition, observation, and interpretation) and their coordination.

- All three elements should be based on modern knowledge of how students learn and how such learning is best measured.
- Considerable time and effort should be devoted to a theory-driven design and validation process before assessments are put into operational use.
When designing new tools for classroom or large-scale use, assessment developers are urged to use the assessment triangle as a guiding framework, as set forth and illustrated in Chapters 5, 6, and 7. As discussed under Recommendation 1 above, a prerequisite for the development of new forms of assessment is that current knowledge derived from research be conveyed to assessment and curriculum developers in ways they can access and use.

A key feature of the approach to assessment development proposed in this report is that the effort should be guided by an explicit, contemporary cognitive model of learning that describes how people represent knowledge and develop competence in the subject domain, along with an interpretation model that is compatible with the cognitive model. Assessment tasks and procedures for evaluating responses should be designed to provide evidence of the characteristics of student understanding identified in the cognitive model of learning. The interpretation model must incorporate this evidence in the assessment results in a way that is consistent with the model of learning. Assessment designers should explore ways of using sets of tasks that work in combination to diagnose student understanding while at the same time maintaining high standards of reliability. The interpretation model must, in turn, reflect consideration of the complexity of such sets of tasks.

An important aspect of assessment validation often overlooked by assessment developers is the collection of evidence that tasks actually tap the intended cognitive content and processes. Starting with hypotheses about the cognitive demands of a task, a variety of research techniques, such as interviews, having students think aloud as they solve problems, and analysis of errors, can be used to explore the mental processes in which examinees actually engage during task performance. Conducting such analyses early in the assessment development process ensures that the assessments do, in fact, measure what they are intended to measure.

**Recommendation 7:** Developers of educational curricula and classroom assessments should create tools that will enable teachers to implement high-quality instructional and assessment practices, consistent with modern understanding of how students learn and how such learning can be measured.

- Assessments and supporting instructional materials should interpret the findings from cognitive research in ways that are useful for teachers.
- Developers are urged to take advantage of opportunities afforded by technology to assess what students are learning at fine levels of detail, with appropriate frequency, and in ways that are tightly integrated with instruction.
The committee believes a synthesis of cognitive and measurement principles has particularly significant potential for the design of high-quality tools for classroom assessment that can inform and improve learning. However, teachers should not be expected to devise on their own all the assessment tasks for students or ways of interpreting responses to those tasks. Some innovative classroom assessments that have emerged from this synthesis and are having a positive impact on learning have been described in preceding chapters. A key to the effectiveness of these tools is that they must be packaged in ways that are practical for use by teachers. As described in Chapter 7, computer and telecommunications technologies offer a rich array of opportunities for providing teachers with sophisticated assessment tools that will allow them to present more complex cognitive tasks, capture and reply to students’ performances, share exemplars of competent performance, engage students in peer and self-reflection, and in the process gain critical information about student competence.

Recommendation 8: Large-scale assessments should sample the broad range of competencies and forms of student understanding that research shows are important aspects of student learning.

- A variety of matrix sampling, curriculum-embedded, and other assessment approaches should be used to cover the breadth of cognitive competencies that are the goals of learning in a domain of the curriculum.
- Large-scale assessment tools and supporting instructional materials should be developed so that clear learning goals and landmark performances along the way to competence are shared with teachers, students, and other education stakeholders. The knowledge and skills to be assessed and the criteria for judging the desired outcomes should be clearly specified and available to all potential examinees and other concerned individuals.
- Assessment developers should pursue new ways of reporting assessment results that convey important differences in performance at various levels of competence in ways that are clear to different users, including educators, parents, and students.

Though further removed from day-to-day instruction than classroom assessments, large-scale assessments also have the potential to support instruction and learning if well designed and appropriately used. Deriving real benefits from the merger of cognitive and measurement theory in large-scale assessment requires finding ways to cover a broad range of competencies.
and capture rich information about the nature of student understanding. Alternatives to the typical on-demand testing scenario—in which every student takes the same test at a specified time under strictly standardized conditions—should be considered to enable the collection of more diverse evidence of student achievement.

Large-scale assessments have an important role to play in providing dependable information for educational decision making by policy makers, school administrators, teachers, and parents. Large-scale assessments can also convey powerful messages about the kinds of learning valued by society and provide worthy goals to pursue. If such assessments are to serve these purposes, however, it is essential that externally set goals for learning be clearly communicated to teachers, students, and other education stakeholders.

Considerable resources should be devoted to producing materials for teachers and students that clearly present both the learning goals and landmark performances along the way to competence. Those performances can then be illustrated with samples of the work of learners at different levels of competence, accompanied by explanations of the aspects of cognitive competence exemplified by the work. These kinds of materials can foster valuable dialogue among teachers, students, and the public about what achievement in a domain of the curriculum looks like. The criteria by which student work will be judged on an assessment should also be made as explicit as possible. Curriculum materials should encourage the use of activities such as peer and self-assessment to help students internalize the criteria for high-quality work and foster metacognitive skills. All of these points are equally true for classroom assessments.

The use of assessments based on cognitive and measurement science will also necessitate different forms of reporting on student progress, both to parents and to administrators. The information gleaned from such assessments is far more nuanced than that obtainable from the assessments commonly used today, and teachers may want to provide more detail in reports to parents about the nature of their children’s understanding. In formulating reports based on new assessments, test developers, teachers, and school administrators should ensure that the reports include the information parents want and can appropriately use to support their children’s learning. Reports on student performance could also provide an important tool to assist administrators in their supervisory roles. Administrators could use such information to see how teachers are gauging their students’ learning and how they are responding to the students’ demonstration of understanding. Such information could help administrators determine where to focus resources for professional development. In general, for the information to be useful and meaningful, it will have to include a profile consisting of multiple elements and not just a single aggregate score.
Teacher Education and Professional Development

Recommendation 9: Instruction in how students learn and how learning can be assessed should be a major component of teacher preservice and professional development programs.

- This training should be linked to actual experience in classrooms in assessing and interpreting the development of student competence.
- To ensure that this occurs, state and national standards for teacher licensure and program accreditation should include specific requirements focused on the proper integration of learning and assessment in teachers’ educational experience.

Research on the integration of cognition and measurement also has major implications for teacher education. Teachers need training to understand how children learn subject matter and how assessment tools and practices can be used to obtain useful information about student competence. Both the initial preparation of teachers and their ongoing professional development can incorporate insights and examples from research on the integration of cognitive and measurement science and equip teachers with knowledge and skills they can use to employ high-quality assessments. At the same time, such learning opportunities can enable teachers to transform their practice in ways that will allow them to profit from those assessments.

If such assessments are to be used effectively, teacher education needs to equip beginning teachers with a deep understanding of many of the approaches students might take toward understanding a particular subject area, as well as ways to guide students at different levels toward understanding (Carpenter, Fennema, and Franke, 1996; Griffin and Case 1997). Teachers also need a much better understanding of the kinds of classroom environments that incorporate such knowledge (NRC, 1999b). Typically, teacher education programs provide very little preparation in assessment (Plake and Impara, 1997). Yet teaching in ways that integrate assessment with curriculum and instruction requires a strong understanding of methods of assessment and the uses of assessment data. This does not mean that all teachers need formal training in psychometrics. However, teachers need to understand how to use tools that can yield valid inferences about student understanding and thinking, as well as methods of interpreting data derived from assessments.

In addition, school administrators need to provide teachers with ample opportunities to continue their learning about assessment throughout their professional practice. Professional development is increasingly seen as a vital element in improving practice, for veteran as well as new teachers.
KNOWING WHAT STUDENTS KNOW

(Cohen and Hill, 1998; Elmore and Burney, 1998). This continued learning should include the development of cognitive models of learning. Teachers' professional development can be made more effective if it is tied closely to the work of teaching (e.g., National Academy of Education, 1999). The “lesson study” in which Japanese teachers engage offers one way to forge this link (Stigler and Hiebert, 1999). In that approach, teachers develop lessons on their own, based on a common curriculum. They try these lessons out in their classrooms and share their findings with fellow teachers. They then modify the lessons and try them again, collecting data as they implement the lessons and again working collaboratively with other teachers to polish them. The resulting lessons are often published and become widely used by teachers throughout the country.

Education Policies

Recommendation 10: Policy makers are urged to recognize the limitations of current assessments, and to support the development of new systems of multiple assessments that would improve their ability to make decisions about education programs and the allocation of resources.

• Important decisions about individuals should not be based on a single test score. Policy makers should instead invest in the development of assessment systems that use multiple measures of student performance, particularly when high stakes are attached to the results.

• Assessments at the classroom and large-scale levels should grow out of a shared knowledge base about the nature of learning. Policy makers should support efforts to achieve such coherence.

• Policy makers should promote the development of assessment systems that measure growth or progress of students and the education system over time and that support multi-level analyses of the influences responsible for such change.

Recommendation 11: The balance of mandates and resources should be shifted from an emphasis on external forms of assessment to an increased emphasis on classroom formative assessment designed to assist learning.

Another arena through which research can influence practice is education policy. This is a particularly powerful arena in the case of assessment. Policy makers currently are putting great stock in large-scale assessments and using them for a variety of purposes. There is a good deal of evidence
that assessments used for policy purposes have had effects on educational practice, not all of them positive (e.g., Herman, 1992; NRC, 1999a; Koretz and Barron, 1998).

Research on the integration of cognition and measurement can affect practice through policy in several ways. Most directly, the research can enhance the assessments used for policy decisions. Furthermore, the decisions of policy makers could be better informed than is the case today by assessments that provide a broader picture of student learning. Since test developers respond to the marketplace, a demand from policy makers for new assessments would likely spur their development.

A range of assessment approaches should be used to provide a variety of evidence to support educational decision making. There is a need for comprehensive systems of assessment consisting of multiple measures, including those that rely on the professional judgments of teachers and that together meet high standards of validity and reliability. Single measures, while useful, are unlikely to tap all the dimensions of competence identified by learning goals. Multiple measures are essential in any system in which high-stakes decisions are made about individuals on the basis of assessment results (NRC, 1999a).

Currently, assessments at the classroom and large-scale levels often convey conflicting goals for learning. As argued in Chapter 6, coherence is needed in the assessment system. A coherent assessment system supports learning for all students. If a state assessment were not designed from the same conceptual base as classroom assessments, the mismatch could undermine the potential for improved learning offered by a system of assessment based on the cognitive and measurement sciences.

To be sure, coherence in an educational system is easier to wish for than to achieve—particularly in an education system with widely dispersed authority such as that of the United States. In many ways, standards-based reform is a step toward achieving some of this coherence. But current content standards are not as useful as they could be. Cognitive research can contribute to the development of next-generation standards that are more effective for guiding curriculum, instruction, and assessment—standards that define not only the content to be learned, but also the ways in which subject matter understanding is acquired and develops. Classroom and large-scale assessments within a coherent system should grow from a shared knowledge base about how students think and learn in a domain of the curriculum. This kind of coherence could help all assessments support common learning goals.

Assessments should be aimed at improving learning by providing information needed by those at all levels of the education system on the aspects of schooling for which they are responsible. If properly conducted, assessments can also serve accountability purposes by providing valuable infor-
mation to teachers and administrators about the progress or growth of the education system over time. The committee refers to this feature as continuity. And if the assessments are instructionally sensitive—that is, if they show the effects of high-quality teaching—they can provide important information about the effectiveness of teaching practices as well (NRC, 1999d).

Developing and implementing a system of multiple assessments would likely be more costly than continuing with the array of tests now being used by states and school districts. Currently, states spend about $330 million for testing (Achieve, 2000). While this sum appears considerable, it represents less than one-tenth of 1 percent of the total amount spent on precollege education (National Center for Education Statistics, 2001). If used properly, the total spending for assessment should not be considered money for tests alone. Funds spent for teachers to score assessments, included in the cost of assessment, also serve an important professional development function. Moreover, spending on assessments that inform instruction represents an investment in teaching and learning, not just in system monitoring. Therefore, policy makers need to invest considerably more in assessment than is currently the case, presuming that the investment is in assessment systems of the type advocated in this report.

Public Opinion and Media Coverage

Recommendation 12: Programs for providing information to the public on the role of assessment in improving learning and on contemporary approaches to assessment should be developed in cooperation with the media. Efforts should be made to foster public understanding of basic principles of appropriate test interpretation and use.

A fourth arena in which research on the integration of cognitive and measurement science can affect practice is through public opinion and the media. Current interest among the public and the news media in testing and test results suggests that public opinion and media coverage can be a powerful arena for change. Information communicated to the public through the media can influence practice in at least two ways. First, the media influence the constituencies responsible for assessment development and practice, including teachers, school administrators, policy makers, and test developers. Perhaps of greater significance is recognition that the more the public is made aware of how assessment practice could be transformed to better serve the goals of learning, the greater will be the support that educators and policy makers have for the kinds of changes proposed in this volume.

Researchers should therefore undertake efforts to communicate with the media what student development toward competence looks like and how it
can best be measured; the media can, in turn, communicate those messages to the public. An attempt should also be made through the media and other avenues for communication with the public to foster understanding of basic principles of appropriate test interpretation and use. Assessment consumers, including the public, should understand that no test is a perfect measure, that more valid decisions are based on multiple indicators, and that the items on a particular assessment are only a sample from the larger domain of knowledge and skill identified as the targets of learning. As part of the design and delivery of such programs, research needs to be conducted on the public’s understanding of critical issues in assessment and the most effective ways to communicate outcomes from educational assessment.

CONCLUDING COMMENTS

As noted at the beginning of this report, educational assessment is an integral part of the quest for improved education. Through assessment, education stakeholders seek to determine how well students are learning and whether students and institutions are progressing toward the goals that have been set for educational systems. The problem is that the vital purposes of informing and improving education through assessment are being served only partially by present assessment practices.

The principles and practices of educational assessment have changed over the last century, but not sufficiently to keep pace with the substantial developments that have accrued in the understanding of learning and its measurement. It is time to harness the scientific knowledge of cognition and measurement to guide the principles and practices of educational assessment. There is already a substantial knowledge base about what better assessment means, what it looks like, and principled ways that can be used to build and use it. That knowledge base needs to be put into widespread practice, as well as continually expanded.

Educators, the public, and particularly parents should not settle for impoverished assessment information. They should be well informed about criteria for meaningful and helpful assessment. To do justice to the students in our schools and to support their learning, we need to recognize that the process of appraising them fairly and effectively requires multiple measures constructed to high standards. Useful and meaningful evidence includes profiling of multiple elements of proficiency, with less emphasis on overall aggregate scores. A central theme of this report is that it is essential to assess diverse aspects of knowledge and competence, including how students understand and explain concepts, reason with what they know, solve problems, are aware of their states of knowing, and can self-regulate their learning and performance.
Achieving these goals requires a strong connection between educational assessments and modern theories of cognition and learning. Without this connection, assessment results provide incomplete, and perhaps misleading, information about what has been learned and appropriate next steps for improvement. Creating better assessments should not be viewed as a luxury, but as a necessity.

Perhaps the greatest challenges to the new science and design of educational assessment relate to disciplinary boundaries and established practices. For instance, there is currently an implicit assumption that one can create good tasks or good assessments and then leave it up to technical people to figure out how to analyze and report the results. Instead, the assessment design process must be a truly multidisciplinary and collaborative activity, with educators, cognitive scientists, subject matter specialists, and psychometricians informing one another during the design process. Other obstacles to pursuing new approaches to assessment stem from existing social structures in which familiar assessment practices are now deeply embedded and thus difficult to change. Professional development and public education are needed to convey how assessment should be designed and how it can be used most effectively in the service of learning.

The investment required to improve educational assessment and further develop the knowledge base to support that effort is substantial. However, this investment in our children and their educational futures is a reasonable one given the public’s legitimate expectation that assessment should both inform and enhance student achievement.