Educational assessment seeks to determine how well students are learning and is an integral part of the quest for improved education. It provides feedback to students, educators, parents, policy makers, and the public about the effectiveness of educational services. With the movement over the past two decades toward setting challenging academic standards and measuring students’ progress in meeting those standards, educational assessment is playing a greater role in decision making than ever before. In turn, education stakeholders are questioning whether current large-scale assessment practices are yielding the most useful kinds of information for informing and improving education. Meanwhile, classroom assessments, which have the potential to enhance instruction and learning, are not being used to their fullest potential.

Advances in the cognitive and measurement sciences make this an opportune time to rethink the fundamental scientific principles and philosophical assumptions serving as the foundations for current approaches to assessment. Advances in the cognitive sciences have broadened the conception of those aspects of learning that are most important to assess, and advances in measurement have expanded the capability to interpret more complex forms of evidence derived from student performance.

The Committee on the Foundations of Assessment, supported by the National Science Foundation, was established to review and synthesize advances in the cognitive sciences and measurement and to explore their implications for improving educational assessment. At the heart of the committee’s work was the critical importance of developing new kinds of educational assessments that better serve the goal of equity. Needed are classroom and large-scale assessments that help all 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.

CONCLUSIONS

The Nature of Assessment and Reasoning from Evidence

This report addresses assessments used in both classroom and large-scale contexts for three broad purposes: to assist learning, to measure individual achievement, and to evaluate programs. The purpose of an assessment determines priorities, and the context of use imposes constraints on the design. Thus it is essential to recognize that one type of assessment does not fit all.

Often a single assessment is used for multiple purposes; in general, however, the more purposes a single assessment aims to serve, the more each purpose will be compromised. For instance, many state tests are used for both individual and program assessment purposes. This is not necessarily a problem, as long as assessment designers and users recognize the compromises and trade-offs such use entails.

Although assessments used in various contexts and for differing purposes often look quite different, they share certain common principles. One such principle is that assessment is always a process of reasoning from evidence. By its very nature, moreover, assessment is imprecise to some degree. Assessment results are only estimates of what a person knows and can do.

Every assessment, regardless of its purpose, rests on three pillars: a model of how students represent knowledge and develop competence in the subject domain, tasks or situations that allow one to observe students’ performance, and an interpretation method for drawing inferences from the performance evidence thus obtained. In the context of large-scale assessment, the interpretation method is usually a statistical model that characterizes expected data patterns, given varying levels of student competence. In less formal classroom assessment, the interpretation is often made by the teacher using an intuitive or qualitative rather than formal statistical model.

Three foundational elements, comprising what is referred to in this report as the “assessment triangle,” underlie all assessments. These three elements—cognition, observation, and interpretation—must be explicitly connected and designed as a coordinated whole. If not, the meaningfulness of inferences drawn from the assessment will be compromised.

The central problem addressed by this report is that most widely used assessments of academic achievement are based on highly restrictive beliefs about learning and competence not fully in keeping with current knowledge about human cognition and learning. Likewise, the observation and interpretation elements underlying most current assessments were created
to fit prior conceptions of learning and need enhancement to support the kinds of inferences people now want to draw about student achievement. A model of cognition and learning should serve as the cornerstone of the assessment design process. This model should be based on the best available understanding of how students represent knowledge and develop competence in the domain.

The model of learning can serve as a unifying element—a nucleus that brings cohesion to curriculum, instruction, and assessment. This cohesive function is a crucial one because educational assessment does not exist in isolation, but must be aligned with curriculum and instruction if it is to support learning.

Finally, aspects of learning that are assessed and emphasized in the classroom should ideally be consistent with (though not necessarily the same as) the aspects of learning targeted by large-scale assessments. In reality, however, these two forms of assessment are often out of alignment. The result can be conflict and frustration for both teachers and learners. Thus there is a need for better alignment among assessments used for different purposes and in different contexts.

Advances in the Sciences of Thinking and Learning

Contemporary theories of learning and knowing emphasize the way knowledge is represented, organized, and processed in the mind. Emphasis is also given to social dimensions of learning, including social and participatory practices that support knowing and understanding. This body of knowledge strongly implies that assessment practices need to move beyond a focus on component skills and discrete bits of knowledge to encompass the more complex aspects of student achievement.

Among the fundamental elements of cognition is the mind’s cognitive architecture, which includes working or short-term memory, a highly limited system, and long-term memory, a virtually limitless store of knowledge. What matters in most situations is how well one can evoke the knowledge stored in long-term memory and use it to reason efficiently about current information and problems. Therefore, within the normal range of cognitive abilities, estimates of how people organize information in long-term memory are likely to be more important than estimates of working memory capacity.

Understanding the contents of long-term memory is especially critical for determining what people know; how they know it; and how they are able to use that knowledge to answer questions, solve problems, and engage in additional learning. While the contents include both general and specific knowledge, much of what one knows is domain- and task-specific and organized into structures known as schemas. Assessments should evaluate what schemas an individual has and under what circumstances he or she regards the infor-
KNOWING WHAT STUDENTS KNOW

This evaluation should include how a person organizes acquired information, encompassing both strategies for problem solving and ways of chunking relevant information into manageable units.

The importance of evaluating knowledge structures comes from research on expertise. **Studies of expert-novice differences in subject domains illuminate critical features of proficiency that should be the targets for assessment.** Experts in a subject domain typically organize factual and procedural knowledge into schemas that support pattern recognition and the rapid retrieval and application of knowledge.

One of the most important aspects of cognition is metacognition—the process of reflecting on and directing one’s own thinking. Metacognition is crucial to effective thinking and problem solving and is one of the hallmarks of expertise in specific areas of knowledge and skill. Experts use metacognitive strategies for monitoring understanding during problem solving and for performing self-correction. **Assessment should therefore attempt to determine whether an individual has good metacognitive skills.**

Not all children learn in the same way and follow the same paths to competence. Children’s problem-solving strategies become more effective over time and with practice, but the growth process is not a simple, uniform progression, nor is there movement directly from erroneous to optimal solution strategies. **Assessments should focus on identifying the specific strategies children are using for problem solving, giving particular consideration to where those strategies fall on a developmental continuum of efficiency and appropriateness for a particular domain of knowledge and skill.**

Children have rich intuitive knowledge of their world that undergoes significant change as they mature. Learning entails the transformation of naive understanding into more complete and accurate comprehension, and assessment can be used as a tool to facilitate this process. To this end, **assessments, especially those conducted in the context of classroom instruction, should focus on making students’ thinking visible to both their teachers and themselves so that instructional strategies can be selected to support an appropriate course for future learning.**

Practice and feedback are critical aspects of the development of skill and expertise. **One of the most important roles for assessment is the provision of timely and informative feedback to students during instruction and learning so that their practice of a skill and its subsequent acquisition will be effective and efficient.**

As a function of context, knowledge frequently develops in a highly contextualized and inflexible form, and often does not transfer very effectively. Transfer depends on the development of an explicit understanding of when to apply what has been learned. **Assessments of academic achievement need to consider carefully the knowledge and skills required to understand and answer a question or solve a problem, including the context in...**
which it is presented, and whether an assessment task or situation is functioning as a test of near, far, or zero transfer.

Much of what humans learn is acquired through discourse and interaction with others. Thus, knowledge is often embedded in particular social and cultural contexts, including the context of the classroom, and it encompasses understandings about the meaning of specific practices such as asking and answering questions. Assessments need to examine how well students engage in communicative practices appropriate to a domain of knowledge and skill, what they understand about those practices, and how well they use the tools appropriate to that domain.

Models of cognition and learning provide a basis for the design and implementation of theory-driven instructional and assessment practices. Such programs and practices already exist and have been used productively in certain curricular areas. However, the vast majority of what is known has yet to be applied to the design of assessments for classroom or external evaluation purposes. Further work is therefore needed on translating what is already known in cognitive science to assessment practice, as well as on developing additional cognitive analyses of domain-specific knowledge and expertise.

Many highly effective tools exist for probing and modeling a person’s knowledge and for examining the contents and contexts of learning. The methods used in cognitive science to design tasks, observe and analyze cognition, and draw inferences about what a person knows are applicable to many of the challenges of designing effective educational assessments.

**Contributions of Measurement and Statistical Modeling to Assessment**

Advances in methods of educational measurement include the development of formal measurement (psychometric) models, which represent a particular form of reasoning from evidence. These models provide explicit, formal rules for integrating the many pieces of information drawn from assessment tasks. Certain kinds of assessment applications require the capabilities of formal statistical models for the interpretation element of the assessment triangle. These tend to be applications with one or more of the following features: high stakes, distant users (i.e., assessment interpreters without day-to-day interaction with the students), complex models of learning, and large volumes of data.

Measurement models currently available can support the kinds of inferences that cognitive science suggests are important to pursue. In particular, it is now possible to characterize student achievement in terms of multiple aspects of proficiency, rather than a single score; chart students’ progress over time, instead of simply measuring performance at a particular point in
time; deal with multiple paths or alternative methods of valued performance; model, monitor, and improve judgments on the basis of informed evaluations; and model performance not only at the level of students, but also at the levels of groups, classes, schools, and states.

Nonetheless, many of the newer models and methods are not widely used because they are not easily understood or packaged in accessible ways for those without a strong technical background. Technology offers the possibility of addressing this shortcoming. For instance, building statistical models into technology-based learning environments for use in classrooms enables teachers to employ more complex tasks, capture and replay students' performances, share exemplars of competent performance, and in the process gain critical information about student competence.

Much hard work remains to focus psychometric model building on the critical features of models of cognition and learning and on observations that reveal meaningful cognitive processes in a particular domain. If anything, the task has become more difficult because an additional step is now required—determining in tandem the inferences that must be drawn, the observations needed, the tasks that will provide them, and the statistical models that will express the necessary patterns most efficiently. Therefore, having a broad array of models available does not mean that the measurement model problem has been solved. The long-standing tradition of leaving scientists, educators, task designers, and psychometricians each to their own realms represents perhaps the most serious barrier to progress.

**Implications of the New Foundations for Assessment Design**

*The design of high-quality classroom and large-scale assessments is a complex process that involves numerous components best characterized as iterative and interdependent, rather than linear and sequential. A design decision made at a later stage can affect one occurring earlier in the process. As a result, assessment developers must often revisit their choices and refine their designs.*

*One of the main features that distinguishes the committee’s proposed approach to assessment design from current approaches is the central role of a model of cognition and learning, as emphasized above. This model may be fine-grained and very elaborate or more coarsely grained, depending on the purpose of the assessment, but it should always be based on empirical studies of learners in a domain. Ideally, the model will also provide a developmental perspective, showing typical ways in which learners progress toward competence.*

*Another essential feature of good assessment design is an interpretation model that fits the model of cognition and learning. Just as sophisticated*
interpretation techniques used with assessment tasks based on impoverished models of learning will produce limited information about student competence, assessments based on a contemporary, detailed understanding of how students learn will not yield all the information they otherwise might if the statistical tools available to interpret the data, or the data themselves, are not sufficient for the task. Observations, which include assessment tasks along with the criteria for evaluating students’ responses, must be carefully designed to elicit the knowledge and cognitive processes that the model of learning suggests are most important for competence in the domain. The interpretation model must incorporate this evidence in the results in a manner consistent with the model of learning.

Validation that tasks tap relevant knowledge and cognitive processes, often lacking in assessment development, is another essential aspect of the development effort. Starting with hypotheses about the cognitive demands of a task, a variety of research techniques, such as interviews, having students think aloud as they work problems, and analysis of errors, can be used to analyze the mental processes of examinees during task performance. Conducting such analyses early in the assessment development process can help ensure that assessments do, in fact, measure what they are intended to measure.

Well-delineated descriptions of learning in the domain are key to being able to communicate effectively about the nature of student performance. Although reporting of results occurs at the end of an assessment cycle, assessments must be designed from the outset to ensure that reporting of the desired types of information will be possible. The ways in which people learn the subject matter, as well as different types or levels of competence, should be displayed and made as recognizable as possible to educators, students, and the public.

Fairness is a key issue in educational assessment. One way of addressing fairness in assessment is to take into account examinees’ histories of instruction—or opportunities to learn the material being tested—when designing assessments and interpreting students’ responses. Ways of drawing such conditional inferences have been tried mainly on a small scale, but hold promise for tackling persistent issues of equity in testing.

Some examples of assessments that approximate the above features already exist. They are illustrative of the new approach to assessment the committee advocates, and they suggest principles for the design of new assessments that can better serve the goals of learning.

Assessment in Practice

Guiding the committee’s work were the premises that (1) something important should be learned from every assessment situation, and (2) the
information gained should ultimately help improve learning. The power of classroom assessment resides in its close connections to instruction and teachers’ knowledge of their students’ instructional histories. Large-scale, standardized assessments can communicate across time and place, but by so constraining the content and timeliness of the message that they often have limited utility in the classroom. Thus the contrast between classroom and large-scale assessments arises from the different purposes they serve and contexts in which they are used. Certain trade-offs are an inescapable aspect of assessment design.

Students will learn more if instruction and assessment are integrally related. In the classroom, providing students with information about particular qualities of their work and about what they can do to improve is crucial for maximizing learning. It is in the context of classroom assessment that theories of cognition and learning can be particularly helpful by providing a picture of intermediary states of student understanding on the pathway from novice to competent performer in a subject domain.

Findings from cognitive research cannot always be translated directly or easily into classroom practice. Most effective are programs that interpret the findings from cognitive research in ways that are useful for teachers. Teachers need theoretical training, as well as practical training and assessment tools, to be able to implement formative assessment effectively in their classrooms.

Large-scale assessments are further removed from instruction, but can still benefit learning if well designed and properly used. Substantially more valid and useful inferences could be drawn from such assessments if the principles set forth in this report were applied during the design process.

Large-scale assessments not only serve as a means for reporting on student achievement, but also reflect aspects of academic competence societies consider worthy of recognition and reward. Thus large-scale assessments can provide worthwhile targets for educators and students to pursue. Whereas teaching directly to the items on a test is not desirable, teaching to the theory of cognition and learning that underlies an assessment can provide positive direction for instruction.

To derive real benefits from the merger of cognitive and measurement theory in large-scale assessment, it will be necessary to devise ways of covering a broad range of competencies and capturing rich information about the nature of student understanding. Indeed, to fully capitalize on the new foundations described in this report will require substantial changes in the way large-scale assessment is approached and relaxation of some of the constraints that currently drive large-scale assessment practices. Alternatives to on-demand, census testing are available. If individual student scores are needed, broader sampling of the domain can be achieved by extracting evidence of student performance from classroom work produced during the
course of instruction. If the primary purpose of the assessment is program evaluation, the constraint of having to produce reliable individual student scores can be relaxed, and population sampling can be useful.

For classroom or large-scale assessment to be effective, students must understand and share the goals for learning. Students learn more when they understand (and even participate in developing) the criteria by which their work will be evaluated, and when they engage in peer and self-assessment during which they apply those criteria. These practices develop students’ metacognitive abilities, which, as emphasized above, are necessary for effective learning.

The current educational assessment environment in the United States assigns much greater value and credibility to external, large-scale assessments of individuals and programs than to classroom assessment designed to assist learning. The investment of money, instructional time, research, and development for large-scale testing far outweighs that for effective classroom assessment. More of the research, development, and training investment must be shifted toward the classroom, where teaching and learning occur.

A vision for the future is that assessments at all levels—from classroom to state—will work together in a system that is comprehensive, coherent, and continuous. In such a system, assessments would provide a variety of evidence to support educational decision making. Assessment at all levels would be linked back to the same underlying model of student learning and would provide indications of student growth over time.

Information Technologies: Opportunities for Advancing Educational Assessment

Information technologies are helping to remove some of the constraints that have limited assessment practice in the past. Assessment tasks no longer need be confined to paper-and-pencil formats, and the entire burden of classroom assessment no longer need fall on the teacher. At the same time, technology will not in and of itself improve educational assessment. Improved methods of assessment require a design process that connects the three elements of the assessment triangle to ensure that the theory of cognition, the observations, and the interpretation process work together to support the intended inferences. Fortunately, there exist multiple examples of technology tools and applications that enhance the linkages among cognition, observation, and interpretation.

Some of the most intriguing applications of technology extend the nature of the problems that can be presented and the knowledge and cognitive processes that can be assessed. By enriching task environments through the use of multimedia, interactivity, and control over the stimulus display, it is pos-
sible to assess a much wider array of cognitive competencies than has here-tofore been feasible. New capabilities enabled by technology include directly assessing problem-solving skills, making visible sequences of actions taken by learners in solving problems, and modeling and simulating complex reasoning tasks. Technology also makes possible data collection on concept organization and other aspects of students' knowledge structures, as well as representations of their participation in discussions and group projects. A significant contribution of technology has been to the design of systems for implementing sophisticated classroom-based formative assessment practices. Technology-based systems have been developed to support individualized instruction by extracting key features of learners' responses, analyzing patterns of correct and incorrect reasoning, and providing rapid and informative feedback to both student and teacher.

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

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

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

**RECOMMENDATIONS FOR RESEARCH, POLICY, AND PRACTICE**

Like groups before us, 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 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. 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: instructional materials, teacher education and professional development, education policies, and public opinion and media coverage. By influencing 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.

The recommendations presented below collectively form a proposed research and development agenda for expanding the knowledge base on the integration of cognition and measurement, and encompass the implications of such a knowledge base for each of the four mediating arenas that directly influence educational practice. Before turning to this agenda, we offer two guidelines for how future work should proceed:

- **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 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 provides a conceptual framework for principled thinking about the assumptions and foundations underlying an assessment.

### Recommendations for Research

**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.

**Recommendation 2:** Funding should be provided for a major program of research, guided by a synthesis of cognitive and measure-
ment 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 models of cognition and 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.

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. This research should also explore how teachers can be assisted in integrating new forms of assessment into their instructional 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. The research should encompass ways in which school structures (e.g., length of time of classes, class size, and opportunity for teachers to work together) affect the feasibility of implementing new types of assessments and their effectiveness.

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, teaching practice, and educational decision making.

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

**Recommendations for Policy and Practice**

**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.

**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 the 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.

**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.
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.

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 also promote the development of assessment systems that measure the growth or progress of students and the education system over time and that support multilevel 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.

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 the basic principles of appropriate test interpretation and use.