The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy’s purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chairman and vice chairman, respectively, of the National Research Council.
COMMITTEE ON THE FOUNDATIONS OF ASSESSMENT

James W. Pellegrino (Co-chair), Peabody College of Education, Vanderbilt University
Robert Glaser (Co-chair), Learning Research and Development Center, University of Pittsburgh
Eva L. Baker, The Center for the Study of Evaluation, University of California, Los Angeles
Paul J. Black, School of Education, King's College, London, England
Christopher L. Dede, Graduate School of Education, Harvard University
Kadriye Ercikan, School of Education, University of British Columbia
Louis M. Gomez, School of Education, Northwestern University
Earl B. Hunt, Department of Psychology, University of Washington
David Klahr, Department of Psychology, Carnegie Mellon University
Richard Lehrer, School of Education, University of Wisconsin
Robert J. Mislevy, School of Education, University of Maryland
Willie Pearson, Jr., Department of Sociology, Wake Forest University
Edward A. Silver, School of Education, University of Michigan
Richard F. Thompson, Department of Psychology, University of Southern California
Richard K. Wagner, Department of Psychology, Florida State University
Mark R. Wilson, School of Education, University of California, Berkeley

Naomi Chudowsky, Study Director
Tina Winters, Research Assistant
M. Jane Phillips, Senior Project Assistant
BOARD ON TESTING AND ASSESSMENT

Eva L. Baker (Chair), The Center for the Study of Evaluation, University of California, Los Angeles
Lorraine McDonnell (Vice Chair), Departments of Political Science and Education, University of California, Santa Barbara
Lauress L. Wise (Vice Chair), Human Resources Research Organization, Alexandria, Virginia
Richard C. Atkinson, President, University of California
Christopher F. Edley, Jr., Harvard Law School
Ronald Ferguson, John F. Kennedy School of Public Policy, Harvard University
Milton D. Hakel, Department of Psychology, Bowling Green State University
Robert M. Hauser, Institute for Research on Poverty, Center for Demography, University of Wisconsin, Madison
Paul W. Holland, Educational Testing Service, Princeton, New Jersey
Daniel M. Koretz, RAND Corporation, Arlington, Virginia
Richard J. Light, Graduate School of Education and John F. Kennedy School of Government, Harvard University
Barbara Means, SRI, International, Menlo Park, California
Andrew C. Porter, Wisconsin Center for Education Research, University of Wisconsin, Madison
Loretta A. Shepard, School of Education, University of Colorado, Boulder
Catherine E. Snow, Graduate School of Education, Harvard University
William L. Taylor, Attorney at Law, Washington, D.C.
William T. Trent, Department of Educational Policy Studies, University of Illinois, Urbana-Champaign
Guadalupe M. Valdes, School of Education, Stanford University
Vicki Vandaveer, The Vandaveer Group, Inc., Houston, Texas
Kenneth I. Wolpin, Department of Economics, University of Pennsylvania

Pasquale J. Devito, Director
Lisa D. Alston, Administrative Associate
Acknowledgments

The work of the Committee on the Foundations of Assessment benefited tremendously from the contributions and good will of many people, and the committee is grateful for their support.

First, we wish to acknowledge the sponsor, the National Science Foundation (NSF). Special thanks go to Larry Suter, who was instrumental in getting the project off the ground and who provided enthusiastic support throughout. We also appreciate the support and valuable input of Elizabeth VanderPutten, Janice Earle, Nora Sabelli, and Eric Hamilton at NSF, as well as Eamonn Kelly, now at George Mason University.

The committee was aided greatly by individuals who participated in a series of information-gathering workshops held in conjunction with several of the committee meetings. We valued the opportunity to hear from a diverse group of researchers and practitioners about the complex issues involved in designing and implementing new forms of assessment.

We wish to make special note of Robbie Case from Stanford University and the Ontario Institute for Studies in Education, who deeply influenced this study. Robbie shared with us his powerful ideas about children’s conceptual development and the implications for assessment and educational equity. Several aspects of his thinking and published work can be found referenced throughout this report. In every respect he was a gentleman and a scholar. His untimely death in 2000 deeply saddened the members of the committee on both a personal and a professional level. His passing represents a major loss for the fields of psychological and educational research.

A number of researchers working at the intersection of cognition and assessment took time to share their work and ideas with the committee, including Drew Gitomer of the Educational Testing Service, Irvin Katz of George Mason University, Jim Minstrell of A.C.T. Systems for Education, Kurt
VanLehn of the Learning Research and Development Center at the University of Pittsburgh, Ken Koedinger of Carnegie Mellon University, Barbara White and John Frederiksen of the University of California at Berkeley, and Jim Greeno of Stanford University. The committee discussed the beliefs and theories of learning underlying some innovative large-scale assessments with Phil Daro of the New Standards Project, Steven Leinwand of the Connecticut State Department of Education, Hugh Burkhardt and Sandy Wilcox of the Mathematics Assessment Resource Service, and Carol Myford of the Educational Testing Service. We also heard from teachers who have used various assessment programs in their classrooms. We thank Guy Mauldin of Science Hill High School, Johnson City, Tennessee; Elizabeth Jones of Walnut Elementary School, Lansing, Michigan; Margaret Davis, Westminster Schools, Atlanta, Georgia; Ramona Muniz, Roosevelt Middle School, San Francisco, California; Cherrie Jones, Alice Carlson Applied Learning Center, Fort Worth, Texas; and Suzanna Loper of the Educational Testing Service, Oakland, CA.

Several individuals discussed special considerations related to disadvantaged students and the design of new forms of assessment. They included Bill Trent of the University of Illinois, Urbana-Champaign, Shirley Malcom of the American Association for the Advancement of Science, Sharon Lewis of the Council of Great City Schools, and Louisa Moats of the National Institute of Child Health and Human Development. Developmental psychologists Susan Goldin-Meadow of the University of Chicago, Robert Siegler of Carnegie Mellon University, and Micki Chi of the Learning Research and Development Center at the University of Pittsburgh discussed research methodologies from their discipline that may have application to educational assessment. A number of researchers helped the committee explore the future role of technology in assessment, including Randy Bennett of the Educational Testing Service, Amy Bruckman of the Georgia Institute of Technology, Walter Kintsch of the University of Colorado, Paul Horwitz of The Concord Consortium, and Gregory Leazer of the University of California at Los Angeles. Lorraine McDonnell of the University of California at Santa Barbara, James Kadamus of the New York State Department of Education, and James Gray of the Dorchester Public Schools in Maryland provided valuable policy perspectives on the prospects for a new science of assessment.

The committee was provided excellent input on advances in statistics and measurement by Steven Raudenbush from the University of Michigan and Brian Junker from Carnegie Mellon University. Their presentations, as well as Brian's commissioned review of statistical methods that are potentially useful for cognitively based assessment, greatly informed our discussions. Linda Steinberg of the Educational Testing Service and Geoff Masters of the Australian Council for Educational Research shared state-of-the-art work on assessment design.

A number of other education researchers provided reactions and syn-
thesizing remarks at the various workshops. They included Bob Linn of the University of Colorado, Rich Shavelson of Stanford University, David Berliner of Arizona State University, Barbara Means of SRI International, Ed Haertel of Stanford University, Goodwin Liu of the U.S. Department of Education, and Nora Sabelli of NSF.

The Board on Testing and Assessment, the unit within the National Research Council (NRC) that launched this study, was instrumental in shaping this project and in providing general guidance and support along the way. Many board members have been mentioned above as participants in the committee's work.

We are especially grateful to several consultants to the project, including Nancy Kober and Robert Rothman, who helped with the writing of this report and provided invaluable assistance in thinking about the organization and presentation of ideas. Rona Briere's skillful editing brought further clarity to our ideas.

Within the NRC, a number of individuals supported the project. Michael Feuer, Director of the Center for Education, conceptualized the project and provided good humor and support along the way. Pasquale DeVito, recently appointed Director of the Board on Testing and Assessment, enthusiastically supported us during the final stages of the project. Patricia Morison offered a great deal of wisdom, advice, and encouragement throughout, and Judy Koenig lent us her substantive knowledge of psychometrics whenever needed. Kirsten Sampson Snyder and Genie Grohman expertly maneuvered us through the NRC review process.

The committee expresses particular gratitude to members of the NRC project staff for contributing their intellectual and organizational skills throughout the study. Three deserve particular recognition. Naomi Chudowsky, the project's study director, was a pleasure to work with and brought incredible talents and expertise to the project. She tirelessly assisted the committee in many ways—serving as a valuable source of information about assessment issues and testing programs; organizing and synthesizing the committee's work; keeping the committee moving forward through its deliberations and the report drafting process; and providing energy, enthusiasm, and exceptional good humor throughout. Her attention to detail while simultaneously helping the committee focus on the bigger picture was a major asset in the creation of the final report. Naomi was assisted by Tina Winters, who provided exceptional research support and adeptly handled preparation of the manuscript. Jane Phillips expertly managed the finances and arranged the meetings for the project, always ensuring that the committee's work proceeded smoothly.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this
independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report: James Greeno, Stanford University; Sharon Griffin, Clark University; Suzanne Lane, University of Pittsburgh; Alan Lesgold, University of Pittsburgh; Marcia C. Linn, University of California, Berkeley; Michael I. Posner, Cornell University; Catherine E. Snow, Harvard University; Norman L. Webb, University of Wisconsin; and Sheldon H. White, Harvard University.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations nor did they see the final draft of the report before its release. The review of this report was overseen by Lauress Wise, Human Resources Research Organization, and Lyle V. Jones, University of North Carolina, Chapel Hill. Appointed by the National Research Council, they were responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Finally, we would like to sincerely thank all of the committee members, who generously contributed their time and intellectual efforts to this project. A study of the scientific foundations of assessment represents an extraordinary challenge, requiring coverage of an exceedingly broad array of complex topics and issues. We were faced with the task of defining the nature of the problem to be studied and solved and then charting a path through a rather ill-defined solution space. Throughout the process, the committee members displayed an extraordinary ability to tolerate ambiguity as we navigated through a vast space of issues and possible answers, at times seemingly without a compass. Simultaneously, they showed a remarkable commitment to learning from each others’ expertise and from the many individuals who shared their knowledge with the group. It has been noted before that the idea of eighteen “experts” collaborating to write a book on any topic, let alone educational assessment, is an absurdity. And yet were it not for the collective expertise, thoughtfulness, and good will of all the committee members, this report and its consensual substantive messages would not have been developed. It has been a professionally stimulating and personally gratifying experience to work with the members of the committee and everyone at the NRC associated with this effort.

Jim Pellegrino, Co-chair
Bob Glaser, Co-chair
In recent years, the National Research Council (NRC), through its Board on Testing and Assessment (BOTA), has explored some of today's most pressing and complex issues in educational assessment. Several NRC committees have examined the role and appropriate uses of assessment in standards-based reform, a movement that is reshaping education throughout the country. For example, committees have studied the impact and uses of tests with high stakes for students, approaches for assessing students with disabilities in a standards-based system, and issues related to proposed voluntary national tests. In the process of carrying out this work, the board and its committees have delved into fundamental questions about educational assessment, such as what its purposes are; which kinds of knowledge and skills should be assessed; how well current assessments, such as the National Assessment of Educational Progress, are fulfilling the various demands placed on them; and which new developments hold promise for improving assessment.

At roughly the same time, other NRC committees have been exploring equally compelling issues related to human cognition and learning. A 1998 report entitled Preventing Reading Difficulties in Young Children consolidates current research findings on how students learn to read and which approaches are most effective for reading instruction. Most recently, the NRC Committee on Developments in the Science of Learning examined findings from cognitive science that have advanced understanding of how people think and learn. The 1999 report of that committee, How People Learn, not only summarizes major changes in conceptions about learning, but also examines the implications of these changes for designing effective teaching and learning environments.

As these multiple committees were progressing with their work, some
NRC staff and members of BOTA decided this would be an ideal time to address a long-standing issue noted by numerous researchers interested in problems of educational assessment: the need to bring together advances in assessment and in the understanding of human learning. Each of these disciplines had produced a body of knowledge that could enrich the other. In fact, some scholars and practitioners were already applying findings from cognitive science in the development of innovative methods of assessment. Although these efforts were generally small-scale or experimental, they pointed to exciting possibilities.

Accordingly, the board proposed that an NRC committee be formed to review advances in the cognitive and measurement sciences, as well as early work done in the intersection between the two disciplines, and to consider the implications for reshaping educational assessment. In one sense, this work would be a natural extension of the conclusions and recommendations of *How People Learn*. In another sense, it would follow through on a desire expressed by many of those involved in the board’s activities to revisit the foundations of assessment—to explore developments in the underlying science and philosophy of assessment that could have significant implications for the long term, but were often glossed over in the short term because of more urgent demands. The National Science Foundation (NSF), recognizing the importance and timeliness of such a study, agreed to sponsor this new NRC effort.

The Committee on the Foundations of Assessment was convened in January 1998 by the NRC with support from NSF. The committee comprised eighteen experts from the fields of cognitive and developmental psychology, neuroscience, testing and measurement, learning technologies, mathematics and science education, and education policy with diverse perspectives on educational assessment.

During its 3-year study, the committee held nine multi-day meetings to conduct its deliberations and five workshops to gather information about promising assessment research and practice. At the workshops, numerous invited presenters shared with the committee members their cutting-edge work on the following topics: (1) assessment practices that are based on cognitive principles and are being successfully implemented in schools and classrooms, (2) new statistical models with promise for use in assessing a broad range of cognitive performances, (3) programs that engage students in self- and peer assessment, (4) innovative technologies for learning and assessment, (5) cognitively based instructional intervention programs, and (6) policy perspectives on new forms of assessment. This report presents the findings and recommendations that resulted from the committee’s deliberations.
Contents

Executive Summary 1

Part I  
Overview and Background

1  Rethinking the Foundations of Assessment 17
2  The Nature of Assessment and Reasoning from Evidence 37

Part II  
The Scientific Foundations of Assessment

Introduction 57

3  Advances in the Sciences of Thinking and Learning 59
4  Contributions of Measurement and Statistical Modeling to Assessment 111
Part III
Assessment Design and Use: Principles, Practices, and Future Directions

Introduction 175

5  Implications of the New Foundations for Assessment Design 177
6  Assessment in Practice 221
7  Information Technologies: Opportunities for Advancing Educational Assessment 261

Part IV
Conclusion

8  Implications and Recommendations for Research, Policy, and Practice 291

References 315

Appendix: Biographical Sketches 349

Index 355
Knowing what Students Know
The Science and Design of Educational Assessment