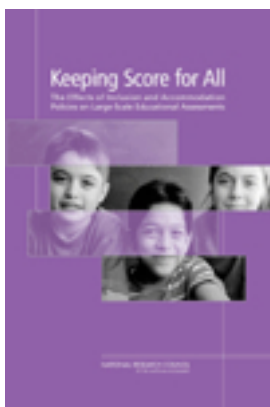


## Free Executive Summary



### **Keeping Score for All: The Effects of Inclusion and Accommodation Policies on Large-Scale Educational Assessment**

Judith Anderson Koenig and Lyle F. Bachman, Editors,  
Committee on Participation of English Language  
Learners and Students with Disabilities in NAEP and  
Other Large-Scale Assessments, National Research  
Council

ISBN: 0-309-09253-1, 150 pages, 6x9, paperback (2004)

This free executive summary is provided by the National Academies as part of our mission to educate the world on issues of science, engineering, and health. If you are interested in reading the full book, please visit us online at <http://www.nap.edu/catalog/11029.html>. You may browse and search the full, authoritative version for free; you may also purchase a print or electronic version of the book. If you have questions or just want more information about the books published by the National Academies Press, please contact our customer service department toll-free at 888-624-8373.

*U.S. public schools are responsible for educating large numbers of English language learners and students with disabilities. This book considers policies for including students with disabilities and English language learners in assessment programs. It also examines the research findings on testing accommodations and their effect on test performance.*

*Keeping Score for All discusses the comparability of states' policies with each other and with the National Assessment of Educational Progress (NAEP) policies and explores the impact of these differences on the interpretations of NAEP results. The book presents a critical review of the research literature and makes suggestions for future research to evaluate the validity of test scores obtained under accommodated conditions. The book concludes by proposing a new framework for conceptualizing accommodations. This framework would be useful both for policymakers, test designers, and practitioners in determining appropriate accommodations for specific assessments and for researchers in planning validity studies.*

**This executive summary plus thousands more available at [www.nap.edu](http://www.nap.edu).**

Copyright © National Academy of Sciences. Permission is granted for this material to be shared for noncommercial, educational purposes, provided that this notice appears on the reproduced materials, the Web address of the online, full authoritative version is retained, and copies are not altered. To disseminate otherwise or to republish requires written permission from the National Academies Press.

# Executive Summary

U.S. public schools are responsible for educating large numbers of students with disabilities and English language learners—some 20 percent of the nation’s 46 million public school students fall into one or both of these categories. Both of these populations have been increasing, and the demand for evidence of their academic progress has also grown. In response to both changing public expectations and legal mandates, the federal government, states, and districts have attempted to include more such students in educational assessments.

Testing these two groups of students, however, poses particular challenges. Many of these students have attributes—such as physical, emotional, or learning disabilities or limited fluency in English—that may prevent them from readily demonstrating what they know or can do on a test. In order to allow these students to demonstrate their knowledge and skills, testing accommodations are used. For the purpose of this report, we have defined testing accommodations by drawing from the definition in the *AERA/APA/NCME Standards for Educational and Psychological Testing* (American Educational Research Association et al., 1999). Our adapted definition is as follows: accommodation is used as the general term for any action taken in response to a determination that an individual’s disability or level of English language development requires a departure from established testing protocol.<sup>1</sup>

---

<sup>1</sup>The actual definition of accommodation in the standards appears in the chapter that deals with the testing of individuals with disabilities and reads as follows: “accommodation is used as the general term for any action taken in response to a determination that an individual’s disability requires a departure from established testing protocol” (American Educational Research Association et al., 1999, p. 101).

The No Child Left Behind Act of 2001 has established the goal for states of including all of their students with disabilities and English language learners in their assessments.<sup>2</sup> At the same time, the sponsors of the National Assessment of Educational Progress (NAEP) hope to increase the participation of these groups of students in NAEP assessments. The use of accommodations provides an important means for increasing inclusion rates for these groups. In identifying appropriate accommodations, policy makers must consider the specific characteristics of the test-takers and the nature of the skills and knowledge (referred to as “constructs”) to be tested. Effective accommodations should not materially alter the nature of the task or the required response, and they should yield scores that are valid indicators of the constructs being assessed. Both state assessment programs and the sponsors of NAEP have set policies regarding the accommodations they will allow. NAEP also has policies for identifying students who cannot meaningfully participate, even with accommodations, and excluding them from the assessment.

However, the existing base of research about the effects of accommodations on test performance and the comparability of scores obtained under standard and accommodated conditions is insufficient to provide empirical support for many of the decisions that must be made regarding the testing of these students. Thus it has been difficult for both state and NAEP officials to make these decisions, and the result has been considerable variation in what is allowed, both from state to state and between NAEP and the state assessments.<sup>3</sup> These kinds of variations in policy, combined with an insufficient research base, create significant impediments to the interpretation of assessment results for both students with disabilities and English language learners.

## STUDY APPROACH

At the request of the U.S. Department of Education, the National Research Council formed the Committee on the Participation of Students with Disabilities and English Language Learners in NAEP and Other Large-Scale Assessments. The charge to the committee was to (1) synthesize research findings about the effects of accommodations on test performance, (2) review the procedures used for making inclusion and accommodation decisions for large-scale assessment programs, and (3) determine the implications of these findings for NAEP inclusion and accommodation policies.

---

<sup>2</sup>The No Child Left Behind Act requires that “not less than 95 percent” of students in each identified subgroup who are enrolled in the school be required to take the assessments used to meet its provisions (P.L. 107-110, Jan. 8, 2002, 115 STAT 1448-1449).

<sup>3</sup>It is important to note that some of this variation can be accounted for by differences in assessment goals, particularly constructs measured, from program to program.

The committee's report addresses three broad areas related to the committee's charge:

1. The policies and practices for the inclusion and provision of accommodations provided for students with disabilities and English language learners that are followed in the National Assessment of Educational Progress and other large-scale assessments conducted by states.
2. The research that has been conducted to date on the effects of accommodations on test performance and the comparability of results from accommodated and standard administrations.
3. The validity of inferences that are made from the results of accommodated assessments.

### **POLICIES AND PROCEDURES FOR INCLUSION AND ACCOMMODATION**

States' policies and procedures for including students with disabilities and English language learners in large-scale assessments have evolved in recent years, and these policies remain in flux as officials strive to refine their procedures for inclusion and accommodation to comply with legislative mandates. These policies and procedures vary widely from state to state, in part because of differences among assessments and assessment systems, and state policies are different from those used for NAEP assessments.

While NAEP's policies are in many cases different from those in place for state assessments, NAEP results are nevertheless affected by state guidelines in two ways. First, NAEP sampling is based on information from the states regarding the characteristics of all of their students. Thus, the samples used to ensure that the population assessed in NAEP is representative of the nation's student population as a whole are dependent on state policies for classifying students as having a disability or being an English language learner, both because states' classification policies and practices vary and because samples from different states may differ in ways that are not explicitly recognized. Second, once NAEP officials identify the sample of students to be included in the assessment, they provide the schools in which those students are enrolled with guidance as to how to administer the assessment. NAEP officials rely on school-level coordinators, who organize the administration of NAEP at schools, to make consistent and logical decisions about which of the students selected in the original sample can meaningfully participate in the assessment. NAEP officials also rely on school coordinators to make decisions about how participating students will be accommodated, on the basis of their individual needs, NAEP's policies, and the accommodations available in that school.

This variability in policies and procedures is important for several reasons. First, NAEP results are reported separately for states so that comparisons can be

made from state to state. If there are differences across states in the characteristics of the sample and in the conditions under which students participate, then the results may not be comparable. Second, national NAEP results are based on the results for each state, so the accuracy of national results is dependent on the consistency of sampling and administration across states. Finally, the policies that govern whether students are included in or excluded from NAEP assessments differ from the policies for inclusion in state assessments. Comparisons of results from a state assessment with those from state NAEP are likely to be affected by these differences.

The accuracy of all data regarding the academic progress of students with disabilities and English language learners is dependent on the uniformity of both the criteria with which students are selected for participation in testing and the administration procedures that are used, including accommodation procedures. In order for the inferences made from assessments of these students to be justifiable, test administration procedures must be uniform. The committee addresses several aspects of this problem with recommendations regarding both policy and research. We address assessment policies first.

The goal of maximizing the participation rates of students with disabilities and English language learners in all testing is widely shared, and is certainly one that the committee endorses. Moreover, the variation in both inclusion and accommodation policies and procedures is too great and has a number of negative effects. The committee therefore makes the following recommendations:<sup>4</sup>

**Recommendation 4-1:** NAEP officials should

- review the criteria for inclusion and accommodation of students with disabilities and English language learners in light of federal guidelines;
- clarify, elaborate, and revise their criteria as needed; and
- standardize the implementation of these criteria at the school level.

**Recommendation 4-2:** NAEP officials should work with state assessment directors to review the policies regarding inclusion and accommodation in NAEP assessments and work toward greater consistency between NAEP and state assessment procedures.

Because NAEP is intended to report on the educational progress of students in the United States, it is important to evaluate the extent to which the results fully represent the student population in each state and in the nation. To evaluate this, the committee reviewed policy materials available from NAEP, NAEP reports,

---

<sup>4</sup>The recommendations are numbered according to the chapter in which they are discussed and the sequence in which they appear in each chapter.

and data available from external data sources (e.g., data reported to Congress under the mandates of the Individuals with Disabilities Education Act, data available from the Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students, and U.S. census data). However, our review revealed a number of issues about which we were concerned. First, our review of NAEP policy materials revealed that there is no clear definition of the target population to which NAEP results are intended to generalize. The policy guidance supplied is not sufficiently specific for making judgments about the extent to which inclusion and exclusion decisions affect the generalizability of the results to the targeted population. Specifically, this guidance does not make clear whether it is intended that all students with disabilities and English language learners should be part of the target population or, if not, which of them are excluded. We therefore recommend that:

**Recommendation 4-3:** NAEP officials should more clearly define the characteristics of the population of students to whom results are intended to generalize. This definition should serve as a guide for decision making and the formulation of regulations regarding inclusion, exclusion, and reporting.

Our review of NAEP reports also revealed that both national and state NAEP reports now indicate the percentages of the NAEP sample that are students with disabilities and English language learners. This is a recent revision to the reports and represents a first step toward making it possible to evaluate the degree to which NAEP samples conform to the definition of the NAEP population. However, the data currently available from state and federal agencies are insufficient to complete the desired comparisons. In the committee's view, it is important to know the extent to which the percentages in the NAEP reports correspond to the percentages of students with disabilities and English language learners reported in other sources. Furthermore, the committee believes that states are undertaking additional efforts at collecting such data, partly in response to the requirements of legislation such as the No Child Left Behind Act of 2001. We encourage all parties (NAEP as well as state and federal agencies) to collect and compile such data so that the desired comparisons can be made. We make two recommendations related to this point:

**Recommendation 4-4:** NAEP officials should evaluate the extent to which their estimates of the percentages of students with disabilities and English language learners in a state are comparable to similar data collected and reported by states, to the extent feasible given the data that are available. Differences should be investigated to determine the causes.

**Recommendation 4-5:** Efforts should be made to improve the availability of data about students with disabilities and English language learners. State-

level data are needed that report the total number of English language learners and students with disabilities by grade level in the state. This information should be compiled in a way that allows comparisons to be made across states and should be made readily accessible.

## RESEARCH REGARDING ACCOMMODATED ASSESSMENTS

The effects of accommodations on test performance have been researched, but the findings that emerge from the existing research are inconclusive. These findings provide little guidance to those who must make decisions about which accommodations are suitable for particular kinds of students participating in particular assessments. What is lacking is research that directly examines the effects of accommodations on the validity of inferences to be made from scores. Overall, existing research does not provide definitive evidence about which procedures would produce the most valid estimates of performance. Moreover, it does not establish that scores for students with disabilities and English language learners obtained under accommodated conditions are as valid as scores for other students obtained under unaccommodated conditions.

For the most part, existing research focuses on comparisons of the scores obtained under standard and accommodated conditions. We conclude that this research design is useful for understanding the effects of accommodations and does provide evidence of differential group performance, but we also conclude that it does not directly address the validity of inferences made from accommodated assessments.

In the committee's judgment, additional types of validity evidence should be collected. Validation studies in which evidence of criterion relatedness is collected have been conducted with the ACT and the SAT; similar studies should be conducted for NAEP and state assessments as well. We acknowledge that identification of appropriate criterion variables is more straightforward in the context of college admissions than in the K-12 context; however, we encourage efforts to identify and obtain reliable data on concurrent measures that can provide evidence of criterion validity for K-12 achievement results, such as grades, teacher ratings, or scores on other assessments of similar constructs. In addition, analyses of test content and test-takers' cognitive processes would provide further insight into the validity of results from accommodated administrations in the K-12 context. We note that NAEP's sponsors have initiated several studies of this kind since our committee began its investigations, and we encourage them to continue in this vein. Specifically, the committee makes the following recommendation:

**Recommendation 5-1:** Research should be conducted that focuses on the validation of inferences based on accommodated assessments of students with disabilities and English language learners. Further research should be guided by a conceptual argument about the way accommodations are intended



to function and the inferences the test results are intended to support. This research should include a variety of approaches and types of evidence, such as analyses of test content, test-takers' cognitive processes, and criterion-related evidence, and other studies deemed appropriate.

**THE VALIDITY OF INFERENCES REGARDING  
ACCOMMODATED ASSESSMENTS**

In an evaluation of a testing program's policies regarding the accommodation of students with disabilities and English language learners, the validity of interpretations of the results should be the primary consideration. A test administered with an accommodation is intended to yield results that are equivalent to the results of a standard administration of the test to a student who has no disability and is fluent in English. However, accommodations can have unintended consequences.

For example, an accommodation might not only allow the student to demonstrate his or her proficiency with regard to the construct being assessed but might also provide that student with an unwarranted advantage over other test-takers. In this case, the resulting score would be an inflated estimate, and hence a less valid indicator, of the test-taker's proficiency.

Thus, determining which accommodation is right for particular circumstances is difficult. The accommodation must at the same time be directly related to the disability or lack of fluency for which it is to compensate and be independent of the constructs on which the student is to be tested. The appropriateness of accommodations might best be understood in terms of a conceptual framework that encompasses both the inferences a test score is designed to support (e.g., the test-taker reads at a particular skill level) and alternative inferences (e.g., the test-taker could not complete the work in the allotted time because of a disability unrelated to his or her skill level on the construct being assessed) that might actually account for the score and therefore impede the collection of the desired information about the test-taker.

Thus the validity of inferences made from the results of any accommodated assessment must be evaluated in terms of the general validation argument for the assessment. That is, there should be a clear definition of the construct the assessment is designed to measure (the targeted skills and knowledge) and the ancillary skills required to demonstrate proficiency on the targeted construct (such as the reading level required to decode the instructions and word problems on an assessment of mathematics skills). Furthermore, the inferences that test designers intend the test results to support should be specified, and evidence in support of claims about how the test results are to be interpreted should be provided.

When accommodations operate as intended, the same kinds of inferences can be drawn from accommodated results as from results based on standard administrations. Only when validation arguments are clearly articulated can the validity



of results from accommodated assessments be evaluated. For this reason, the committee examined the available documentation of the constructs to be assessed and the validity evidence laid out for NAEP assessments.

The committee concludes that the validation argument for NAEP in general is not as well articulated as it should be. NAEP officials have not explicitly described the kinds of inferences they believe their data should support, and we found insufficient evidence to support the validity of inferences made from accommodated NAEP scores. While arguments in support of the validity of accommodated administrations of NAEP are discussed in some NAEP materials, more extensive and systematic investigation of the validity of inferences made from these scores is needed. At the same time, as has been noted, existing research does not provide definitive evidence about which procedures will, in general, produce the most valid estimates of performance for students with disabilities and English language learners.

The committee presents a model for evaluating the validity of inferences made from accommodated assessments, based in part on the evidence-centered design approach that has been developed by Hansen, Mislevy, and Steinberg (Hansen and Steinberg, 2004; Hansen et al., 2003; see also Mislevy et al., 2003). This model offers a means of disentangling the potential explanations for observed performance on an assessment and using this analysis to discern the effects of accommodations on the validity of inferences to be based on the observed performance. This approach provides a first step in laying out validity arguments to be investigated through empirical research.

We make three recommendations regarding validity research on accommodations. Although these recommendations are specific to NAEP, we strongly urge the sponsors of state and other large-scale assessment programs to consider them as well.

**Recommendation 6-1:** NAEP officials should identify the inferences that they intend should be made from its assessment results and clearly articulate the validation arguments in support of those inferences.

**Recommendation 6-2:** NAEP officials should embark on a research agenda that is guided by the claims and counterclaims for intended uses of results in the validation argument they have articulated. This research should apply a variety of approaches and types of evidence, such as analyses of test content, test-takers' cognitive processes, criterion-related evidence, and other studies deemed appropriate.

**Recommendation 6-3:** NAEP officials should conduct empirical research to specifically evaluate the extent to which the validation argument that underlies each NAEP assessment and the inferences the assessment was designed to support are affected by the use of particular accommodations.

## CONCLUSION

The difficulties related to assessing students with disabilities and English language learners are not new, but the consequences of relying on scores whose accuracy cannot be ensured have become even greater because of the provisions of the No Child Left Behind Act of 2001. Under that legislation, states are responsible for tracking the academic progress of the students with disabilities and English language learners in every school. The consequences for a school of failing to ensure that these students make progress every year toward ambitious targets of performance are serious. However, regardless of that legislation or any modifications that may be made to it, the validity of test-based inferences made about the performance of students with disabilities and English language learners will be critical for those who seek to understand the academic progress of these students, as well as for those who make policies that affect them.

Under the present circumstances, the need for tests results in which users can have justifiable confidence is, if not more critical, at least more immediate. The No Child Left Behind Act requires schools and jurisdictions to take their legal obligations to assess English language learners and students with disabilities more seriously than many have done in the past. While the committee considers this renewed attention to the needs of both groups of students an important development in the effort to close persistent achievement gaps, the goal cannot be met without accurate data. Credible assessment results can play a crucial role in revealing not only where schools are failing these students, but also where they are succeeding with them. Thus it is essential that evidence of the validity of assessment results be thoroughly investigated to be sure that these results can provide useful information regarding students with disabilities and English language learners for schools, local jurisdictions, and the nation.

# Keeping Score for All

The Effects of Inclusion and Accommodation Policies on Large-Scale Educational Assessments

Committee on Participation of English Language Learners and  
Students with Disabilities in NAEP and Other Large-Scale Assessments

Judith Anderson Koenig and Lyle F. Bachman, Editors

Board on Testing and Assessment  
Center for Education  
Division of Behavioral and Social Sciences and Education

NATIONAL RESEARCH COUNCIL  
*OF THE NATIONAL ACADEMIES*

THE NATIONAL ACADEMIES PRESS  
Washington, D.C.  
**[www.nap.edu](http://www.nap.edu)**

**THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001**

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract/Grant No. R215U990016 between the National Academy of Sciences and the U.S. Department of Education. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

**Library of Congress Cataloging-in-Publication Data**

National Research Council (U.S.). Committee on Participation of English Language Learners and Students with Disabilities in NAEP and Other Large-Scale Assessments.

Keeping score for all : the effects of inclusion and accommodation policies on large-scale educational assessments / Committee on Participation of English Language Learners and Students with Disabilities in NAEP and Other Large-Scale Assessments ; Judith Anderson Koenig and Lyle F. Bachman, editors.

p. cm.

Includes bibliographical references.

ISBN 0-309-09253-1 (pbk.) — ISBN 0-309-53317-1 (pdf)

1. Educational tests and measurements—United States—Evaluation. 2. Inclusive education—United States—Evaluation. 3. Students with disabilities—Education—United States. 4. Limited English-proficient students—Education—United States. 5. National Assessment of Educational Progress (Project) I. Koenig, Judith A. II. Bachman, Lyle F., 1944- III. Title. LB3060.83.N295 2004 371.26'4—dc22

2004014364

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, <http://www.nap.edu>.

Printed in the United States of America

Copyright 2004 by the National Academy of Sciences. All rights reserved.

Suggested citation: National Research Council. (2004). *Keeping Score for All: The Effects of Inclusion and Accommodation Policies on Large-Scale Educational Assessments*. Committee on Participation of English Language Learners and Students with Disabilities in NAEP and Other Large-Scale Assessments. Judith A. Koenig and Lyle F. Bachman, Editors. Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

# THE NATIONAL ACADEMIES

## *Advisers to the Nation on Science, Engineering, and Medicine*

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. Harvey V. Fineberg 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 chair and vice chair, respectively, of the National Research Council.

**[www.national-academies.org](http://www.national-academies.org)**



**COMMITTEE ON PARTICIPATION OF ENGLISH LANGUAGE  
LEARNERS AND STUDENTS WITH DISABILITIES IN NAEP AND  
OTHER LARGE-SCALE ASSESSMENTS**

**Lyle F. Bachman** (*Chair\**), Department of Applied Linguistics, University of California, Los Angeles  
**Jonathan G. Dings**, Boulder Valley School District, Boulder, CO  
**Judy L. Elliott**, Long Beach Unified School District, San Pedro, CA  
**Margaret J. McLaughlin**, Institute for the Study of Exceptional Children and Youth, University of Maryland, College Park  
**Mark D. Reckase** (*Chair\*\**), Measurement and Quantitative Methods, Michigan State University  
**Lourdes C. Rovira**, Miami-Dade County Public Schools, Miami, FL  
**María Medina Seidner**, Texas Education Agency, Austin, TX  
**Rebecca Zwick**, Gevirtz Graduate School of Education, University of California, Santa Barbara

**Judith A. Koenig**, *Study Director*  
**Alexandra Beatty**, *Senior Program Officer*  
**Michael DeCarmine**, *Senior Program Assistant*

---

\*Beginning November 2003.  
\*\*Stepped down from the Committee in November 2003, when he was appointed to the National Assessment Governing Board.



**BOARD ON TESTING AND ASSESSMENT**

- Eva L. Baker** (*Chair*), Center for the Study of Evaluation, University of California, Los Angeles
- Lorraine McDonnell** (*Vice Chair*), Department of Political Science, University of California, Santa Barbara
- Laurens L. Wise** (*Vice Chair*), Human Resources Research Organization, Alexandria, VA
- Christopher F. Edley, Jr.**, Harvard Law School
- Emerson J. Elliott**, Independent Consultant, Arlington, VA
- Milton D. Hakel**, Department of Psychology, Bowling Green State University, OH
- Robert M. Hauser**, Institute for Research on Poverty, Center for Demography, University of Wisconsin–Madison
- Paul W. Holland**, Educational Testing Service, Princeton, NJ
- Daniel M. Koretz**, Graduate School of Education, Harvard University
- Edward P. Lazear**, Graduate School of Business, Stanford University
- Richard Light**, Graduate School of Education, John F. Kennedy School of Government, Harvard University
- Robert J. Mislevy**, Department of Measurement, Statistics, and Evaluation, University of Maryland, College Park
- James W. Pellegrino**, Department of Psychology, University of Illinois, Chicago
- Lorrie Shepard**, Department of Education, University of Colorado, Boulder
- 
- Stuart W. Elliott**, *Director*
- Lisa D. Alston**, *Administrative Associate*

# Preface

The Committee on the Participation of Students with Disabilities and English Language Learners in NAEP and Other Large-Scale Assessments was formed under the auspices of the National Research Council (NRC) with the support of the National Center for Education Statistics (NCES). Its report addresses critical issues in the assessment of students with disabilities and English language learners, an issue that has come to the forefront of conversations about the effects of the No Child Left Behind Act of 2001. The NRC’s Board on Testing and Assessment (BOTA) has focused considerable attention during its ten-year history on the challenges and questions presented by the need to include these students in assessment and accountability programs. At present these students make up approximately 20 percent of the nation’s 46 million public school students, and while the responsibility for monitoring their progress is not new either to the National Assessment of Educational Progress (NAEP) or to states and districts, the provisions of the No Child Left Behind Act of 2001 have made that responsibility more public, more complex, and more urgent.

A November 2001 BOTA workshop, also sponsored by NCES, focused on the reporting and interpretation of test results for students with disabilities and English language learners who receive test accommodations. Discussions at the workshop made clear that several key issues merited more in-depth examination. The first set of concerns pertained to the way decisions regarding both the inclusion of students with special needs in assessments and the identification of appropriate accommodations for them are made. It was clear from the workshop not only that there is considerable variability across states and districts in the way these decisions are made but also that this variability can affect NAEP results in

significant ways. The second set of concerns pertained to the research that has been conducted on the effects of accommodations on test performance. The 2001 workshop stimulated considerable discussion about the conclusions that could be drawn from the existing literature base and about appropriate research approaches for evaluating the effects of accommodations on test performance.

Therefore, at the request of NCES, the Committee on the Participation of Students with Disabilities and English Language Learners in NAEP and Other Large-Scale Assessments was formed and given a three-part charge, to (1) synthesize research findings about the effects of accommodations on test performance, (2) review the procedures used for making inclusion and accommodation decisions for large-scale assessment programs, and (3) determine the implications of these findings for NAEP inclusions and accommodation policies. The work of this committee was intended to build on the discussions from the 2001 workshop (and other earlier NRC projects).

Thus, the report from the 2001 workshop served as the starting point for the committee's work. In addition, researcher Stephen Sireci was commissioned to conduct a review and critique of the research literature on the effects of accommodations on test performance. This literature review focused on empirical studies conducted between 1990 and 2003 and was commissioned and completed in time to be ready for the committee's first meeting in March 2003. The workshop report and this literature review served as background information for the committee as it began its in-depth examination of the relevant policies and practices in effect around the country and the state of the research in this vital area.

The resulting report is designed to set the committee's findings and recommendations in the context of current policy and practice with regard to the inclusion and accommodation of students with disabilities and English language learners. In this report, the committee discusses the meaning of scores from accommodated assessments and the kinds of evidence that are needed to support inferences made from these scores. It is BOTA's hope that this report will be of use to both the officials who oversee NAEP and those who oversee state and local assessments as they work to make their assessments as inclusive as possible, and to make them yield results that accurately reflect the knowledge and skills of all students.

The committee is very grateful to the many individuals who have helped with this project from its inception. It takes particular note of the contribution of Mark Reckase of Michigan State University, who served as its chair through most of its work, until his appointment to the National Assessment Governing Board required him to step down. The committee also sincerely appreciates Lyle Bachman's willingness to assume the responsibility of chair for the completion of the project.

The committee extends its heartfelt thanks to Peggy Carr of NCES for her interest in this important topic and her willingness to fund the project. The committee also thanks Arnold Goldstein of NCES for his constant and prompt support

and quick answers to all of the committee's questions. The committee also appreciates the assistance of Debra Hollinger Martinez, formerly of NCES, who provided valuable materials and information. Nancy Caldwell of Westat provided information about NAEP's sampling procedures and responded to committee questions about NAEP's administrative procedures for accommodating students with disabilities and English language learners. Jim Carlson of NAGB provided the committee with information on NAEP policies on accommodation as well as on research on the effects of accommodations on NAEP results. Mary Crovo of NAGB spoke with the committee about the constructs assessed by NAEP, and the committee is grateful to all of these busy officials for their time and assistance.

Many others assisted the committee as well. Researchers Martha Thurlow of the National Center on Education Outcomes and Charlene Rivera of the Center for Equity and Excellence provided the committee with background about states' policies; John Olsen of the Council of Chief State School Officers (CCSSO) made a presentation about the results of the CCSSO's annual survey of states' practices. Presentations by Wayne Camara of the College Board and Robert Ziomek of ACT, Inc., on their research on the effects of accommodations on admissions test performance were particularly informative. The committee is grateful to Ed Haertel of Stanford University for providing an advance copy of his paper on evidentiary arguments and the comparability of scores from standard and nonstandard test administrations. Eric Hansen of Educational Testing Service briefed the committee on his work on the evidence-centered design approach and prepared a commissioned paper that was very helpful to the committee. The committee is also indebted to Stephen Sireci of the University of Massachusetts, Amherst and his colleagues Stanley Scarpati and Shuhong Li for their work on the commissioned review and critique of the literature.

The committee also wishes to thank the NRC staff who have supported the project. Study director Judith Koenig has offered leadership and support in countless ways and has guided the committee through some quite complicated territory. The committee also thanks Pasquale DeVito for his initial work on the project, and Stuart Elliott and Patricia Morison, who read and made valuable comments on several versions of the report. Michael DeCarmine's calm and able administrative assistance throughout the project is much appreciated as well. The committee is also grateful for Alexandra Beatty's expert writing ability. She provided invaluable assistance in drafting portions of the report and editing it so that it read with one voice. NRC editor Chris McShane provided valuable editing and smoothing in the final stages as well. Finally, the committee is indebted to Kirsten Sampson Snyder for ably guiding the report through the NRC review process.

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 RRC. 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 ensuring 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. I thank the following individuals for their review of this report: Jamal Abedi, Technical Projects, Center for the Study of Evaluation, University of California, Los Angeles; Susan A. Agruso, Office of Assessment, Charlotte-Mecklenburg School District, Charlotte, NC; Richard Duran, Gevirtz Graduate School of Education, University of California, Santa Barbara; Shelley Loving-Ryder, Assistant Superintendent for Assessment and Reporting, Virginia Department of Education; Diana Pullin, Education Law and Public Policy, Boston College; and Martha Thurlow, National Center on Educational Outcomes, University of Minnesota, Minneapolis.

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 Lyle V. Jones, L.L. Thurstone Psychometric Laboratory, University of North Carolina. Appointed by the National Research Council, he was 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.

Eva L. Baker  
*Chair, Board on Testing and Assessment*

# Contents

Executive Summary	1
1 Introduction	10
2 Characteristics of the Students, the Assessments, and Commonly Used Accommodations	18
3 Participation in NAEP and Other Large-Scale Assessments	41
4 Factors That Affect the Accuracy of NAEP’s Estimates of Achievement	62
5 Available Research on the Effects of Accommodations on Validity	85
6 Articulating Validation Arguments	103
7 Implications of the Committee’s Conclusions and Recommendations	123
References	126
Appendix: Biographical Sketches of Committee Members and Staff	133





# Keeping Score for All

The Effects of Inclusion and Accommodation  
Policies on Large-Scale Educational Assessments

