NATIONAL CENTER FOR EDUCATION STATISTICS

Statistical Analysis Report — February 2000

Increasing the Participation of Special Needs Students in NAEP

A Report on 1996 NAEP Research Activities

U.S. Department of Education
Office of Educational Research and Improvement — NCES 2000-473

What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history, geography, and other fields. By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individual students and their families.

NAEP is a congressionally mandated project of the National Center for Education Statistics (NCES), within the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations. NAEP reports directly to the Commissioner, who is responsible for providing continuing reviews, including validation studies and solicitation of public comment, on NAEP's conduct and usefulness.

In 1988, Congress established the National Assessment Governing Board (NAGB) to formulate policy guidelines for NAEP. The Board is responsible for selecting the subject areas to be assessed from among those included in the National Education Goals; for setting appropriate student performance levels; for developing assessment objectives and test specifications through a national consensus approach; for designing the assessment methodology; for developing guidelines for reporting and disseminating NAEP results; for developing standards and procedures for interstate, regional, and national comparisons; for determining the appropriateness of test items and ensuring they are free from bias; and for taking actions to improve the form and use of the National Assessment.

The National Assessment Governing Board

NATIONAL CENTER FOR EDUCATION STATISTICS

Statistical Analysis Report **February 2000**

Increasing the Participation of Special Needs Students in NAEP

A Report on 1996 NAEP Research Activities

John Mazzeo
James E. Carlson
Kristin E. Voelkl
Anthony D. Lutkus

in collaboration with
John J. Ferris
Bruce A. Kaplan
Edward Kulick
Steven P. Isham

U.S. Department of Education
Office of Educational Research and Improvement **NCES 2000-473**

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
C. Kent McGuire
Assistant Secretary

National Center for Education Statistics
Gary W. Phillips
Acting Commissioner

Assessment Division
Peggy G. Carr
Associate Commissioner

February 2000

SUGGESTED CITATION
U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics. *Increasing the Participation of Special Needs Students in NAEP: A Report on 1996 NAEP Research Activities,* NCES 2000–473, by J. Mazzeo, J. E. Carlson, K. E. Voelkl, & A. D. Lutkus. Washington, DC: 2000.

FOR MORE INFORMATION
Content contact:
Arnold A. Goldstein
202–219–1741

To obtain single copies of this report, while supplies last, or ordering information on other U.S. Department of Education products, call toll free 1–877– 4ED PUBS (877–433–7827), or write:

Education Publications Center (ED Pubs)
U.S. Department of Education
P.O. Box 1398
Jessup, MD 20794–1398

TTY/TDD 1–877–576–7734
FAX 301–470–1244

Online ordering via the Internet: *http://www.ed.gov/pubs/edpubs.html*
Copies also are available in alternate formats upon request.
This report also is available on the World Wide Web: *http://nces.ed.gov/nationsreportcard/*

The work upon which this publication is based was performed for the National Center for Education Statistics, Office of Educational Research and Improvement, by Educational Testing Service.

Table of Contents

Tables

Executive Summary

Increasing the Participation of Special Needs Students in NAEP: A Report on 1996 NAEP Research Activities

This study grew out of concerns about the underrepresentation of students with special needs in the National Assessment of Educational Progress (NAEP) assessments. The term "special needs students" is sometimes used to include both students with disabilities and students who are limited English proficient (LEP). In the 1996 NAEP assessment samples, 10 percent of fourth graders, 9 percent of eighth graders, and 5 percent of twelfth graders were identified by their schools as students with disabilities. In the same assessment year, 4 percent of fourth graders and 2 percent of eighth and twelfth graders were identified by their schools as students with limited English proficiency.[1] Schools participating in NAEP have been permitted to exclude individuals they identify as special needs students from the assessment, in accordance with criteria provided by the program at that time. In fact, at least half of all special needs students were excluded from NAEP assessments in 1992 and 1994. This exclusion has raised concerns that some special needs students who could be meaningfully assessed are being excluded from NAEP. Moreover, there is an additional concern that variations across locales in exclusion practices may introduce biases in NAEP results.

In recent years, a number of policy, legislative, civil rights, and technical considerations have caused the NAEP program to look more closely at its administration and assessment procedures and to consider changes that can increase participation among students with disabilities and LEP students. Based on previous studies,[2] as well as recommendations from various offices in the U.S. Department of Education, program procedures have been modified with the aim of increasing participation among special needs students. Modifications were made in two areas.[3] First, inclusion criteria for the NAEP 1996 assessment were revised with the intention of making them clearer, more inclusive, and more likely to be applied consistently across jurisdictions participating in the state assessment program. Second, for the first time in NAEP, a variety of assessment accommodations were offered to 1) students with disabilities

---

[1] Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). *NAEP 1996 mathematics report card for the nation and the states.* (p. 67). Washington, DC: National Center for Education Statistics.

[2] National Academy of Education. (1993). *The trial state assessment: Prospects and realities; the third report of the national academy of education panel on the evaluation of the NAEP 1992 trial state assessment*, Stanford, CA: Author.

[3] Olson, J.F., & Goldstein, A.A. (1996). Increasing the inclusion of students with disabilities and limited English proficient students in NAEP. *Focus on NAEP. 2*(1). Washington, DC: National Center for Education Statistics.

whose Individualized Education Plan (IEP) specified such accommodations for testing; and 2) LEP students who, in the opinion of their instructors, required an accommodation in order to take the assessment in English.

This report presents in-depth analyses of the effects on inclusion rates of the above efforts to increase the participation of special needs students in NAEP. It also contains an analysis of selected technical characteristics of the assessment results and a review of descriptive results of the background characteristics and educational experiences of students with disabilities and LEP students who participated in the NAEP 1996 national assessments in mathematics and science.

In particular, data are presented on:

- the possible effect on the NAEP proficiency scales of including greater percentages of special needs students;
- the comparability of results from nonstandard administrations (i.e., administrations in which accommodations were allowed) to results obtained under standard conditions; and
- the effect of nonstandard administrations on NAEP's capacity to provide accurate comparisons of trends over time.

In addition, it is important to be clear on what this report does not contain:

- This report does not provide an in-depth examination of the performance on NAEP of students with disabilities and LEP students.
- The relatively small sample sizes obtained in the study did not allow disaggregation of students with disabilities and LEP students in many of the statistical analyses that dealt with the effects on NAEP scales.
- This report does not separate students with disabilities from LEP students in the Differential Item Functioning (DIF) analyses.
- This report does not look at performance results or inclusion rates for students with disabilities and LEP students by state.

A experiment was designed for the 1996 assessments in mathematics and science, which permitted analysis of data relevant to the issues above. In addition, a questionnaire was included that was designed to obtain information on student background and educational experiences. The questionnaire was to be completed for all sampled students with disabilities and for all sampled LEP students. The design of the NAEP 1996 assessment included three distinct national samples of schools. In the first of these school samples (denoted S1), the assessment was conducted using the same inclusion criteria used during the 1990 and 1992 NAEP assessments in mathematics and science. In the second school sample (denoted S2), revised inclusion criteria were used. No assessment accommodations or adaptations were offered to students in S1 or S2 schools. In the third sample (denoted S3), the assessment was conducted using inclusion criteria that were identical to those used in S2 schools. The S3 sample was distinguished, however, by the availability of a variety of assessment accommodations and adaptations. To ensure sufficient amounts of data for planned analyses, students with disabilities and LEP students were oversampled in national S2 and S3 schools, and all students in S3 who received an accommodation at a given grade were administered the same NAEP assessment booklet.

Major Findings

Technical Characteristics of Results

The findings of the current research on technical characteristics of the assessment results based on the combined data from all special-needs students include the following:

- For two of the three grades in science there is some evidence to suggest that test results obtained using accommodations and adaptations cannot be fit with the same Item Response Theory (IRT)[4] model as results obtained under standard administration conditions. The evidence for the mathematics assessment was less conclusive. Because small sample sizes necessitated the combination of students with disabilities and limited English proficient students for IRT and scaling analyses, it is not yet clear whether future NAEP reports will need to report these categories of students separately. A future report using larger samples (combined state data) from the 1998 Reading assessment should shed further light on this question.
- Despite the finding above, the inclusion of data from nonstandard administrations had no discernable effect on aggregate NAEP scaling results in mathematics and science at any of the three grades. Differences in test characteristic curves and test information curves plotted with and without the inclusion of such data differed no more than would be expected due to sampling variability.
- Proficiency means were estimated for the NAEP mathematics and science scales, with and without the inclusion of students with accommodations in testing at each of grades 4, 8, and 12. There were no significant differences in the overall means or in the means for significant subgroups at any of the three grades.
- The results reported here suggest that the procedural changes being considered would not significantly affect the NAEP scale score results. If so, it may be possible for the NAEP program to achieve its joint goals of increasing inclusion while maintaining trend lines. However, additional research is necessary to determine the generality of these findings across content areas and over time, as state policies and procedures with respect to inclusion evolve.

Data from background questionnaires did allow separate analyses for SD and LEP students pertaining to background characteristics, educational experiences, and inclusion rates. Major findings for these analyses are summarized below.

---

[4] IRT analyses provides a common scale on which performance can be compared across groups such as those defined by grade and characteristics, including gender and race/ethnicity.

Students with Disabilities

Background Characteristics, Educational Experiences and Inclusion Rates

- Learning disability was by far the most frequently reported category for students with disabilities, with close to three of four students so identified at each of the three grades.
- About half of the students at each grade was described as having mild disabilities. The remaining half at each grade was almost all categorized with moderate to severe disabilities. Very few students receiving special educational services at schools participating in NAEP (1 percent at grades four and eight, and 3 percent at grade twelve) were judged to have profound disabilities.
- Regardless of grade level, about half of all students with disabilities were mainstreamed in academic subjects at least 80 percent of the time.
- In reading/language arts, half or fewer of the students with disabilities received instruction that was at grade level. In mathematics and science, the situation was slightly better at the two lower grades. More than half of the grade four and grade eight students with disabilities received grade-level instruction, and over 70 percent of these students received grade-level instruction in science.
- Almost all students who received instruction that was at or above grade level received the same curriculum content as their nondisabled peers. In contrast, fewer than half of those students with disabilities who received below grade-level instruction was taught the same curriculum content as their nondisabled peers.
- In all three grades, more than 75 percent of students with disabilities were judged by school personnel to be performing below grade level in reading/language arts. Reported performance levels in mathematics and science were somewhat higher than those in reading/language arts at grade four.
- Across the three grades, respondents reported that 42 to 44 percent of students with disabilities received some form of accommodation or adaptation in testing.

Inclusion Rates

- Comparison of questionnaire results with actual participation rates from the 1996 mathematics assessment suggest that: 1) increases in the percentages of students with disabilities participating in NAEP are not likely to result solely from revisions to inclusion criteria; and, 2) a further expansion of accommodations or adaptations permitted by NAEP, or a change in NAEP guidelines as to who is eligible for special testing conditions, could result in further small increases in inclusion percentages.
- Most exclusion decisions were made on the basis of what is stated in the IEP, and relatively few exclusion decisions were made on the basis of a judgment of severe cognitive impairment, absent corroborating direction from the IEP. However, results also suggest that, for substantial percentages of excluded students, neither

determination by the IEP team nor the presence of cognitive impairments was given as reason for exclusion.

- Some students who do not regularly receive accommodations or adaptations were offered them in NAEP and others who should not have been tested were, in fact, included. These results suggest that incorrect decisions regarding inclusion or testing condition may have been made or that incorrect questionnaire data may have been provided.

Students with Limited English Proficiency

Background Characteristics and Educational Experiences

- The largest proportion of LEP students spoke Spanish as their native language (74 percent at grade four, 72 percent at grade eight, and 54 percent at grade twelve). The most frequently encountered other languages were Vietnamese, Hmong, Chinese, Russian, and Pacific Island languages.

- Forty-four percent of grade-four LEP students, 47 percent of grade-eight LEP students, and 65 percent of grade-twelve LEP students had received academic instruction primarily in English for three or more years.

- At grades eight and twelve, few students received native-language instruction in academic areas. At grade four, the percentages of LEP students who received native- language instruction in reading/language arts, mathematics, and science were 22, 27, and 26 percent, respectively.

- Among LEP students receiving English-language instruction, the majority received instruction at grade level at all three grades.

- The vast majority of LEP students at all three grades (87 percent of grade-four LEP students, 80 percent of grade-eight LEP students and 81 percent of grade-twelve LEP students) received some special academic instruction in English or in their native language. At grades eight and twelve, such special instruction appears to have been predominantly in English.

- Although most LEP students were receiving English-language *instruction* on grade-level, a significant percentage were judged to be *performing* below grade level in English. In reading/language arts, where one might expect the impact of limited language proficiency to be most pronounced, 70 percent of grade-four and 62 percent of grade-eight LEP students were judged by school personnel as performing below grade level in English; at grade twelve, 50 percent were so judged. In science, the percentages reported performing below grade level ranged from 30 at grade twelve to 44 at grade eight. In mathematics, the percentages ranged from 33 percent (at grade twelve) to 46 percent (at grade eight).

- Respondents indicated that 37 percent of grade-four LEP students, 27 percent of grade-eight LEP students, and 22 percent of grade-twelve LEP students used accommodations and adaptations in achievement testing in their schools.

Inclusion Rates

- The operational criteria used in NAEP from 1990 to 1996 indicated that LEP students enrolled in schools where English is the primary language of instruction for two or more years were to be included in the assessment. At least 85 percent of fourth-grade students, 67 percent of eighth-grade students, and 83 percent of twelfth-grade students had been enrolled for two or more years in schools where English was the primary language. Historically, NAEP inclusion rates for LEP students have been below the ideal minimums suggested by questionnaire results.
- As was the case for students with disabilities, comparisons of questionnaire results with assessment inclusion rates for LEP students suggest that: 1) increases in the percentage of LEP students are not likely to result solely from revisions to inclusion criteria that do not also involve the provision of accommodations; and 2) further modest improvements in inclusion might still be possible if the list of permitted accommodations and adaptations can be expanded.
- Analyses of inclusion rates by the length of time students were enrolled in schools where English is the primary language of instruction provided some evidence that, when implemented without the provision of accommodations and adaptations, the revised criteria actually resulted in less inclusion among LEP students than did the original criteria. This evidence was strongest at grade four.
- Under the revised criteria, all students receiving academic instruction in English for three or more years were to be included in NAEP. Analyses based on questionnaire responses as to the number of years students were receiving academic instruction in English indicated that this ideal was not quite achieved. Inclusion rates among students with three or more years of academic instruction in English were high, but total inclusion was not achieved, even where accommodations and adaptations were provided.
- Some LEP students who do not usually receive accommodations in testing were apparently provided accommodations in the NAEP assessment. The percentages of LEP students in this category were small (10, 6, and 5 percent in grades four, eight, and twelve, respectively).
- Questionnaire results suggest that the procedural modifications made to NAEP had their primary impact on inclusion rates at grades four and eight among students who would be tested in their native language if this accommodation were available. Participation rates for these students were higher when accommodations were available.

Chapter 1

Introduction

The National Assessment of Educational Progress (NAEP) is the only nationally representative and continuing assessment of what students in the United States know and can do in various academic subjects. Authorized by Congress, directed by the National Center for Education Statistics (NCES) of the U.S. Department of Education, and provided with policy guidance by the National Assessment Governing Board (NAGB), NAEP has become a valuable tool in tracking progress toward the National Education Goals.[1] Because NAEP's purpose is to report on what the nation's students know and can do, it is important that NAEP results represent the educational attainments of ***all*** students. This includes the educational attainments of special needs students – students with disabilities and limited English proficient (LEP) students. While the intent of NAEP has always been to include such students in its assessments to the fullest degree possible, the administration of the survey has always resulted in some exclusion of students with disabilities and students with limited English proficiency.

During the past three years, NCES has sponsored a number of research projects aimed at increasing the participation of special needs students and providing reliable and valid measurement of their levels of academic attainment.[2] A key part of this research effort involved collecting data on students with disabilities and LEP students, and investigating ways of increasing the participation of these students in the NAEP 1996 assessments in mathematics and science. Data collection was accomplished using a questionnaire designed to obtain detailed information on the background characteristics and educational experiences of each special needs student. An experiment was designed to investigate the impact of two procedural modifications – a revision of NAEP inclusion criteria for special needs students and the provision of accommodations and adaptations in NAEP testing – on the percentages of these students included in NAEP and the overall technical quality of the assessment results.

---

[1] Executive Office of the President. (1990). *National goals for education.* Washington, DC: U.S. Government Printing Office. Goals 2000: Educate America Act, H.R. 1804, 103d Cong., 2nd Sess. (1994).

[2] Olson, J.F., & Goldstein, A.A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progress.* (NCES Publication No. 97-482). Washington, DC: National Center for Education Statistics.

Some initial results on the impact of these modifications on inclusion rates were presented in the NAEP 1996 report cards in mathematics and science.[3] This report presents additional and more detailed findings from NAEP's 1996 research activities involving students with disabilities and LEP students. Specifically, it contains descriptive information on the background characteristics and educational experiences of special needs students, a more detailed examination of the impact of the procedural modifications on inclusion rates for these students, and analyses directed at key psychometric issues that are by products of the efforts to increase inclusion.

Students With Disabilities

During the 1995-96 school year, the Office of Special Education Programs (OSEP) reported that approximately 5 million students aged 6-21 were served under the Individuals with Disabilities Education Act, Part B.[4] The 1996 NAEP assessment in mathematics reported that about 10 percent of the NAEP national population of grade-4 students was identified as having disabilities, and that nine and five percent were so classified at grades 8 and 12, respectively.[5] Recent data show these percentages to be increasing. For example, the 1994-95 OSEP figure represents an overall increase of about 3 percentage points since the 1993-94 school year. Similarly, about 11 percent of students sampled for the 1994 NAEP Trial State Assessment in Reading had an Individualized Education Plan (IEP) related to disability, compared to nine percent in 1992.[6] (Note that throughout this report "IEP" refers to Individualized Education Plan, as used in the 1996 and 1997 Annual Reports to Congress on the implementation of IDEA. After 1998, IEP is often used as an abbreviation for Individualized Education Program).

Historically, students with disabilities have been protected under the law against discrimination and also provided educational assistance. Under the Individuals with Disabilities Act (IDEA), students with diabilities who require special education services are afforded the right to a free and appropriate public education that must be tailored to the individual's learning needs. That is, Individualized Education Plans (IEPs) are set up for students with disabilities to specify educational objectives and services. In 1997, amendments were made to IDEA which, among other things, require states to include students with disabilities in statewide testing, to report the number of students with disabilities assessed, to offer accommodations or alternate testing situations to facilitate their inclusion, and to report in

---

[3] Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). *NAEP 1996 mathematics report card for the nation and the states*. Washington, DC: National Center for Education Statistics.

O'Sullivan, C.Y., Reese, C.M., & Mazzeo, J. (1997). *NAEP 1996 science report card for the nation and the states.* Washington, DC: National Center for Education Statistics.

[4] Office of Special Education Programs. (1997). *Nineteenth annual report to Congress on the implementation of the individuals with disabilities education act.* Washington, DC: U.S. Department of Education.

[5] Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). *NAEP 1996 mathematics report card for the nation and the states*. Washington, DC: National Center for Education Statistics.

[6] The National Academy of Education. (1996). *Quality and utility: The 1994 trial state assessment in reading. The fourth report of the national academy of education panel on the evaluation of the NAEP trial state assessment: 1994 trial state assessment in reading.* Stanford, CA: Author.

similar fashion the performance of all students. However, the 1996 NAEP assessments were conducted prior to the passing of these amendments. Nearly 10 percent of school-age children have disabilites and qualify for special education services under IDEA.[7]

While the IDEA specifically addresses educational issues related to students with disabilities, other federal civil rights laws also protect persons with disabilities from discrimination on the basis of disability. Section 504 of the Rehabilitation Act of 1973 addresses disability issues for any organization receiving receiving federal funding. Title II and Title III of the Americans with Disabilities Act of 1990 applies to all services, programs, and activities provided by or made available by state or local government entities, including public schools. These laws were designed to ensure that individuals with disabilities are not excluded from participation in or denied the benefits of public services, programs, or activities by reason of their disability. Students with disabilities who do not require special education services but do require disability-related aids and services to participate in public programs are covered by Section 504 and the American Disabilities Act, but not by the IDEA.[8]

Students with disabilities are a heterogeneous group. According to the IDEA, students can be classified with respect to their primary disabilities into one of 12 categories: autism, deaf-blindness, hearing impairments, mental retardation, multiple disabilities, orthopedic impairment, other health impairment, serious emotional disturbance, specific learning disability, speech or language impairment, traumatic brain injury, and visual impairment. Data from the 1995-1996 school year indicate that most students with disabilities fall into one of four primary disability categories: specific learning disabilities, speech or language impairments, mental retardation, or serious emotional disturbances (about half of all students who qualify for special education have specific learning disabilities as their primary disability category). The degree of disability can range from mild to severe. Some students with disabilities are fully integrated into the generalized education classroom and participate in the general curriculum, while others receive specialized curricula and individualized instruction.

The lack of uniformity and reliability in identifying which students have disabilities has complicated the decision to include them in educational assessments. Poor reliability stems from problems with overlapping diagnostic categories, differences between teachers' degree of tolerance for student differences, and differences within and between districts on screening and placement practices.[9] In addition, the categories used to identify disabilities are not uniform across states, so that individual students may be classified differently from state to state. Differences in classification practices across states have continued to present challenges to the NAEP program in that Westat (the NAEP's sampling and data collection contractor) depends on the schools to identify which students have disabilities. Most state guidelines for participation in assessments rely on the student's Individualized Education Plan (IEP), although some variability exists on using parental input, student placement, and availability of alternate assessments for guidance.[10]

---

[7] McDonnell, L.M., McLaughlin, M.J., & Morison, P. (Eds.). (1997). *Educating one & all: Students with disabilities and standards based reform*. Washington, DC: National Academy of Science. National Research Council.

[8] U.S. Department of Education. (1997). *Nineteenth annual report to Congress on the implementation of the Individuals with Disabilities Education Act.* Washington, DC: Author.

[9] Ibid.

[10] Olson, J.F., & Goldstein, A.A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progress.* (NCES Publication No. 97-482). Washington, DC: National Center for Education Statistics.

Limited English Proficient Students

The term "limited English proficient" (LEP) is often used to refer to language minority students whose proficiency in English has not yet developed sufficiently to allow them to participate fully in an English-only instructional environment.[11] As with students with disabilities, LEP students are a heterogeneous group with respect to their demographic and instructional experiences. Limited English proficient students are protected under both state and federal laws, which assert that they are eligible for special services if they cannot meaningfully and equitably participate in an English-only school environment.

As with students with disabilities, the number of limited English proficient students in the United States is increasing. The 1990 U.S. Census estimated that approximately 14 percent of the U.S. student population lived in homes where a language other than English was spoken.[12] According to the Office of Bilingual Education and Minority Language Affairs (OBEMLA), in 1995 there were approximately 3.2 million limited English proficient students in the United States. This was an increase of 4.8 percent from the previous year.[13]

The 1996 NAEP mathematics assessment reported that about 4, 2, and 2 percent of students at grades 4, 8, and 12, respectively, were limited English proficient.[14] Similarly, about six percent of the fourth-grade students sampled for the 1994 NAEP Trial State Assessment in Reading was limited English proficient.[15] In general, greater proportions of LEP students are found in the lower elementary years of schooling (K-4), with decreasing percentages observed as the grade levels increase.

Including Students With Disabilities and Limited English Proficient Students in NAEP

Mandated by Congress, NAEP examines the educational accomplishments of fourth-, eighth-, and twelfth-grade students in the United States and monitors changes in those accomplishments. NAEP samples include students with disabilities and LEP students in percentages similar to those in which they are represented in the general school population. It is the intention of NAEP that its results be representative of all students in the nation. To this end, the NAEP program has for some time sought to provide uniform guidelines for inclusion. However, because of the voluntary nature of the program, the implementation of the guidelines

---

[11] Alternate terms such as "language minority students," "bilingual," and "English language learner" have also been used to describe LEP students.

[12] August, D., & Hakuta, K. (Eds.). (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Research Council, Institute of Medicine, Commission on Behavioral and Social Sciences and Education.

[13] Olson, J.F., & Goldstein. A.A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progress.* (NCES Publication No. 97-482, p. 57). Washington, DC: National Center for Education Statistics.

[14] Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). *NAEP 1996 mathematics report card for the nation and the states*. Washington, DC: National Center for Education Statistics.

[15] The National Academy of Education. (1996). *Quality and utility: The 1994 trial state assessment in reading. The fourth report of the national academy of education panel on the evaluation of the NAEP trial state assessment: 1994 trial state assessment in reading.* Stanford, CA: Author.

depends on decisions made by local school personnel. These decisions vary from district to district and from state to state.

While the intention of NAEP is to provide a reliable and valid portrait of proficiency for all students in the nation at a given age or grade, many schools excluded students with disabilities and LEP students from NAEP. For example, about half or more of the students identified as having disabilities or who were limited English proficient were excluded from NAEP assessments in 1992 and 1994. Among the students with disabilities and LEP students sampled for NAEP in 1996, more than half were assessed, but 41 percent of students with disabilities and 44 percent of LEP students were not.[16]

Among students with disabilities, customary reasons for exclusion include:

- the IEP specified that the student should not be assessed;
- participation in a different curriculum and enrollment in special classes; and
- beliefs among school staff that certain students are unable to participate meaningfully in the assessment.

Tables 3.5 and 3.8 in chapter 3 of this report present data pertinent to the relative importance of such reasons in inclusion decisions. Among LEP students, a customary reason for exclusion is the inability of the student to take the test in English. Tables 5.3 and 5.8 in chapter 5 of this report present data pertinent to the need for test instruments in languages other than English. Among students with disabilities and LEP students, some students might not have participated in NAEP because the assessment was unable to provide the accommodations or adaptations that would have made their inclusion possible. Again, data presented in chapters 3 and 5 provide information on the increases in inclusion rates that are evident when such accommodations and adaptations can be made available.

Recently, the National Academy of Education (NAE) conducted two evaluations of NAEP to inform the inclusiveness issue.[17] The reports revealed that while many students with disabilities and LEP students had been excluded from previous NAEP assessments, many of them were actually capable of participating in the assessment. This was found to be true particularly if certain types of adaptations and accommodations were made available to the students. For example, the panel for the NAE's evaluation of the 1994 Trial State Assessment (TSA) in fourth-grade reading reported that as many as 85 percent of students with disabilities could read well enough to be included in NAEP. Further, among the sample of LEP students in this evaluation (LEP students attending schools where English was the primary language of instruction for at least two years), more than half had been excluded from the Trial State

---

[16] Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). *NAEP 1996 mathematics report card for the nation and the states*. Washington, DC: National Center for Education Statistics.

[17] National Academy of Education. (1993). *The trial state assessment: Prospects and realities: The third report of the national academy of education panel on the evaluation of the NAEP 1992 trial state assessment*. Stanford, CA: Author.

National Academy of Education. (1996). *Quality and utility: The 1994 trial state assessment in reading. The fourth report of the national academy of education panel on the evaluation of the NAEP trial state assessment: 1994 trial state assessment in reading.* Stanford, CA: Author.

Assessment even though more than three-quarters of the sample had been in schools where English was the primary language of instruction for more than four years.

Many of the students who were excluded from the 1994 TSA were deemed capable of participating, if accommodations had been available. For example, teachers of both students with disabilities and LEP students in the 1994 TSA were likely to recommend testing accommodations for high percentages of their students. If these recommendations had been followed, only 20 percent of the sample of students with disabilities would have been assessed without accommodations, compared to the 56 percent who actually were assessed. Similarly for LEP students, only half as many students in the study would have been assessed under the 1994 TSA conditions. The panel noted that these findings concur with those of the 1995 NAEP field test, in which accommodations were provided for a large proportion of students who would have been included without accommodations. Despite these findings, the knowledge base that currently exists on the inclusion of students with disabilities and students with limited English proficiency in large-scale assessments is quite incomplete.[18]

The issue of inclusion is intertwined with recent standards-based reform movements in this country. These movements advocate setting high academic standards that define the knowledge and skills to be taught by teachers and learned by students, measuring what students are expected to know, and determining whether equal learning opportunities have been given to all students.

Popular convictions about equity in setting academic standards and about the assessment of all students in meeting these standards coincide with attention on inclusiveness in NAEP. In recent years, a number of policy, legislative civil rights, and technical considerations have caused NAEP to look more closely at its administration and assessment procedures and to consider changes that can increase participation among students with disabilities or LEP students.[19] During the 1994 assessment cycle, the National Center for Education Statistics (NCES) and the National Assessment Governing Board (NAGB) began to reexamine policies and practices related to the exclusion of students with disabilities and LEP students from NAEP. Based on previous studies[20, 21] as well as recommendations from various offices in the U.S. Department of Education, program procedures have been modified to increase the participation of these groups of students. Modifications were made in two areas.[22] First, inclusion criteria were revised with the intention of making them clearer, more inclusive, and more likely to be applied consistently across jurisdictions participating in the state assessment program. Second, a variety of assessment accommodations and adaptations was offered to students with disabilities whose IEPs specified such accommodations for testing and

---

[18] Olson, J.F., & Goldstein, A.A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progress.* (NCES Publication No. 97-482). Washington, DC: National Center for Education Statistics.

[19] Olson, J.F., & Goldstein, A.A. (1996). Increasing the inclusion of students with disabilities and limited English proficient students in NAEP. *Focus on NAEP. 2*(1). Washington, DC: National Center for Education Statistics.

[20] National Academy of Education. (1993). *The trial state assessment: Prospects and realities: The third report of the national academy of education panel on the evaluation of the NAEP 1992 trial state assessment*. Stanford, CA: Author.

[21] Ysseldyke, J.E., Thurlow, M.L, McGrew, K.S., & Vanderwood, M. (1994). *Making decisions about the inclusion of students with disabilities in statewide assessments* (Synthesis Report No. 13). Minneapolis: University of Minnesota, National Center for Education Outcomes.

[22] Olson, J.F., & Goldstein, A.A. (1997). Increasing the inclusion of students with disabilities and limited English proficient students in NAEP. *Focus on NAEP. 2*(1). Washington, DC: National Center for Education Statistics.

to LEP students who were, in the opinion of their instructors, unable to take the assessment in English. These guidelines were tried out in the NAEP 1995 field test and later implemented in the 1996 assessment.

Findings From the NAEP 1995 Field Test. Inclusion procedures were revised for the NAEP 1995 field test so that more students would be included in the assessment and the inclusion policies were more consistent. Both the revised inclusion criteria and the provision of accommodations were implemented. The field test was composed of two special studies: 1) an LEP study to examine the impact of changing the inclusion criteria to be more consistent with current state practices and to determine whether a bilingual or Spanish-language version of the mathematics test could be administered and validly scaled; and 2) a study to examine the impact of changing the inclusion criteria for students with disabilities and to determine whether it was reasonable and valid to provide accommodations such as large print or braille test booklets.

The procedures and results of the 1995 field test are detailed in a separate report.[23] In general, the findings showed that the new inclusion procedures could be implemented on a national level in the 1996 assessment and that the provision of accommodations and adaptations allowed many students with disabilities and LEP students to participate who might not have otherwise. The results of the field test also raised questions about the effect accommodations have on trend measurements, including issues of student participation and achievement. The results from the field test were incorporated into plans for the 1996 assessment.

Design for the NAEP 1996 Mathematics and Science Assessments

The 1996 national and state NAEP mathematics and science assessments were conducted in a manner that ensured the reporting of valid trend results. Samples of students were assessed using materials and administration procedures consistent with those used for the 1990 and 1992 assessments. In addition to these core assessment activities, the 1996 assessment included supplemental samples of schools and students. The supplemental samples were designed to allow the program to study the feasibility and impact of increasing the numbers of LEP students and students with disabilities who are included in NAEP and are assessed in an appropriate manner (e.g., with accommodations called for in their IEP's or required by the student to meaningfully participate). Specifically, revised inclusion rules were implemented in one sample; assessment accommodations and adaptations were permitted in another.
In addition, the NAEP program introduced a new background questionnaire, to be filled out for all sampled students with disabilities and LEP students, designed to obtain descriptive information on student background characteristics and educational experiences.

[23] Anderson, N.E., Jenkins, F.F., & Miller, K.E. (1996). *NAEP inclusion criteria and testing accommodations: Findings from the NAEP 1995 field test in mathematics.* Princeton, NJ: Educational Testing Service.

The Three Samples. The design of the NAEP 1996 mathematics assessment required three distinct national samples of schools and two distinct samples of schools within each jurisdiction that participated in the state assessment program. Note that the school was the sampling unit, and each school was randomly assigned to one of three possible samples. In the first of these school samples (denoted S1), the assessment was conducted using the same inclusion criteria as used during the 1990 and 1992 NAEP assessments in mathematics and science. In the second of the school samples (denoted S2), revised inclusion criteria were used. No assessment accommodations or adaptations were offered to students in S1 or S2 schools. In 1996, S1 and S2 school samples were selected at all three grades for the national assessment, and at grades 4 and 8 for the jurisdictions participating in state NAEP assessments. In the third school sample (denoted S3), the assessment was conducted using the same inclusion criteria as were used in the S2 sample. The S3 sample was distinguished, however, by the availability of a variety of assessment accommodations and adaptations. In 1996, S3 samples were selected at all three grades of the national assessment. Table 1.1 summarizes the 1996 NAEP national school sampling conditions.

To ensure sufficient amounts of data for planned analyses, students with disabilities and LEP students were oversampled in national S2 and S3 schools, and all students in S3 who received an accommodation at a given grade were administered the same NAEP assessment booklet. A decision was made to gain greater experience with this modification to existing procedures in the context of the smaller scale, more controlled conditions of the national assessment before extending the procedure to large-scale implementation, as would be required for the state NAEP assessments. Moreover, the decision to go forward with offering accommodations and adaptations in operational NAEP was made after states had agreed to participate. Therefore, providing accommodations and adaptations in state NAEP was not part of the participating states' agreement.

Table 1.1 — Sampling design of the NAEP 1996 national mathematics and science assessments for students with disabilities and students with limited English proficiency

THE NATION'S REPORT CARD naep

School Sample	Mathematics	Science
S1	Used 1900-1996 operational inclusion criteria	Not applicable
S2	Used revised inclusion criteria (greater inclusion)	Used revised inclusion criteria (greater inclusion)
S3	Used revised inclusion criteria plus testing accommodations and adaptations	Used revised inclusion criteria plus testing accommodations and adaptations

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics and Science Assessments.

Data from S1 and a portion of S2 (students without IEPs or equivalent plans) were combined and analyzed as the reporting sample appropriate for national and state comparisons with previous NAEP assessments. By comparing results obtained from S1 to those from S2, in this report, NAEP is able to assess the effects of changing inclusion criteria on inclusion rates and assessment results. Similarly, by comparing results obtained from S2 with those from S3, the program is able to assess the effects of providing accommodations and adaptations. Finally, by comparing results from S1 and S3, the program is able to assess the effects of jointly changing the inclusion criteria and providing accommodations and adaptations.

Overall, the designs for the mathematics and science assessments were similar. Since the 1996 NAEP science assessment was based on a new framework, the national NAEP science assessment did not require or include an S1 sample of schools. In order to report state-level trend results in mathematics, an S1 sample of schools was required in each jurisdiction. In order to avoid increasing the number of schools required for the state NAEP samples, these same schools also conducted the science assessment under S1 conditions. Thus S1 samples in states afforded comparison of the old versus the new inclusion criteria in both mathematics and science.

For the science assessment, S2 samples were drawn for all three grade levels in the national assessment and for each of the jurisdictions participating in the grade 8 state assessment. As with the mathematics assessment, separate S3 samples were not obtained for the state assessments. Therefore, the national assessment in science was based on two samples (an S2 and an S3 sample).

Revised Inclusion Criteria and Provision of Accommodations. Revised inclusion criteria for NAEP were implemented on an experimental basis in the S2 and S3 samples for the 1996 assessment. The revision had the following four goals:

1. Increase inclusion rates for students with disabilities.
2. Bring NAEP inclusion rules for LEP students more in line with those used in state testing programs.
3. Allow for more consistent inclusion decisions across states and jurisdictions.
4. Ensure that inclusion decisions were related to the subject-matter instruction given to the student rather than less relevant considerations.

Original inclusion criteria (used in S1) provided a basis for determining whether students could be excluded from the assessment. Based on the S1 criteria (i.e., the criteria used in NAEP's mathematics assessments in 1990 and 1992), students with disabilities could be excluded only if they were mainstreamed in academic subjects less than 50 percent of the time **and/or** judged to be incapable of participating meaningfully in the assessment. LEP students could be excluded if they were native speakers of a language other than English **and** enrolled in a school where English is the primary language of instruction for less than two years, **and** judged to be incapable of taking part in the assessment.

The 1996 guidelines used in S2 were revised to emphasize criteria for the inclusion rather than exclusion of students with disabilities and LEP students. Although the original criteria did instruct school staff to include students when in doubt, the revised criteria were designed to communicate more clearly a presumption of inclusion except under special circumstances. Students with IEPs were to be included in the NAEP assessment unless they fell into one of the following cases:

1. The school's IEP team determined that the student could not participate; **or**
2. The student's cognitive functioning was so severely impaired that she or he could not participate; **or**
3. The student's IEP required that the student be tested with an accommodation or adaptation, and that the student could not demonstrate his or her knowledge without that accommodation.

Under the revised criteria, all LEP students receiving academic instruction in English for three years or more were to be included in the assessment. Those LEP students receiving instruction in English for less than three years were to be included unless school staff judged them to be incapable of participating in the assessment in English.

In the S3 sample, the revised criteria were used, and various accommodations and adaptations were made available. Adaptations generally refer to changes made specifically to the test format, such as large-print or braille test books. Accommodations are usually associated with changes in the testing environment and administration process. These include, for example, allowing extended time to take the test or being tested individually rather than in a group. NAEP attempted to assess students with disabilities under conditions identical to those under which they normally participate in large-scale assessments. To the extent possible, NAEP offered S3 students the assessment accommodations that were specified in their IEP or equivalent document. For example, if a student's IEP specified that he or she could only be assessed with extended assessment time, NAEP provided this accommodation. Thus, students whose IEPs required accommodations or adaptations were included in NAEP if the program was able to offer them the prescribed accommodation.

In general, most accommodations that schools routinely provided for their own testing were allowed in S3. These permitted accommodations included the following:

- One-on-one testing
- Small group testing
- Extended time
- Oral reading of directions
- Signing of directions
- Use of magnifying equipment
- Use of an amanuensis (an aide who manually transcribes student responses onto the answer sheets)

NAEP also developed a braille version of the mathematics instrument at grade 8 and a large-print version at grades 4 and 8. These modified-format booklets were made available to students who normally would have been assessed using braille or large-print materials.

It should be noted that students assessed under one of the special conditions typically received some combination of accommodations and adaptations. For example, students assessed in small groups (as opposed to standard NAEP sessions of roughly 30 students) usually received extended time and had directions and/or assessment questions read aloud as needed. In one-on-one administrations, students often received assistance in recording answers, had directions and questions read aloud, and were afforded extra time.

NAEP goals and plans regarding LEP students were somewhat different. As with students with disabilities, the new inclusion criteria emphasized inclusion rather than exclusion, and LEP students were eligible for any of the accommodations previously listed. However, field test experience had suggested that many LEP students did not have formal plans that specified assessment accommodations. Because the majority of these students are native Spanish speakers, a translation of the instrument seemed to offer an opportunity to include many students who had been excluded in the past. Therefore, in addition to the accommodations listed above, LEP students at grades 4 and 8 were offered a bilingual version of the mathematics assessment, which displayed Spanish and English versions of questions on facing pages. In S3 this version was administered to LEP students whose teachers believed that the student could only participate in NAEP if given this version, or that the student could best show his or her mathematical abilities by working with this instrument. Students who took this booklet were typically assessed in a small-group setting. In addition, a Spanish/English glossary of scientific terms used in the science assessment was produced. This glossary was made available to students in all three grades who, when tested, normally make use of such a document or who typically receive related accommodations (such as a bilingual dictionary). Use of the glossary was permitted in standard NAEP testing sessions, as well as in small-group and one-on-one testing situations. Students using the bilingual booklet and glossary were typically given extra time.

The SD/LEP Questionnaire. SD/LEP questionnaires were distributed at each participating school. A knowledgeable school staff member was asked to fill out the questionnaire for each student with a disability and each LEP student in the NAEP sample, regardless of whether or not the student was judged capable of participating in the NAEP assessment. The questionnaire contained a 30-question section on students with disabilities to be completed for all students in the sample who were eligible for special education services under the IDEA or as a result of section 504 of the Rehabilitation Act. The questions in this section were designed to provide information on the nature and severity of the student's disabilities, type of educational experiences, and the circumstances of his or her participation in large-scale assessments like NAEP. The questionnaire also contained a 36-question section to be completed for all students in the sample who were identified by their schools as LEP. The questions in this section were designed primarily to provide information on their educational experiences and the circumstances of their participation in large-scale assessments like NAEP. A copy of the full questionnaire is provided in appendix B.

Findings From the NAEP 1996 Assessment

Preliminary answers to several important research questions were reported in the 1996 NAEP Report Cards in Mathematics and Science.

- The introduction of the revised inclusion criteria, without the provision of accommodations, had little effect on the percentage of the total population that was assessed in NAEP at either the national or state level.
- Likewise, the introduction of the revised inclusion criteria, without the provision of accommodations, had, at most, a limited effect on the percentage of students with disabilities or LEP students who were assessed in NAEP at either the national or state level.
- The provision of accommodations and adaptations in the mathematics assessment clearly increased participation rates for students with disabilities and LEP students at grades 4 and 8. When accommodations or adaptations were available, more than 70 percent of both of these groups were assessed at each of these two grades. These numbers are substantially higher than the program has achieved in past assessments, where accommodations and adaptations were not offered. On the other hand, providing accommodations at grade 12 had little effect on inclusion rates for either group.
- The provision of accommodations and adaptations did not increase the numbers of LEP students included in the science assessment. This may be partially explained by the absence of a bilingual booklet and the lack of use of the Spanish/English glossary. Among students with disabilities, the provision of accommodations resulted in significantly higher inclusion rates at grade 4 only.
- A portion of the population of students with disabilities was assessed with accommodations or adaptations when these were available. However, there is evidence that some students who used accommodations could have been assessed without them. A similar pattern of results was not evident among LEP students.

Issues in Need of Further Study

The previously reported findings discussed above provided a good starting point for understanding the impact of the procedural modifications being implemented in NAEP. However, more in-depth study was undertaken for this report. For example, this report makes use of the extensive information provided by responses to the SD/LEP questionnaire on student background, educational experiences, and the conditions under which special needs students participate in large-scale assessments such as NAEP. In addition to providing a rich descriptive database, the SD/LEP questionnaire data provide an opportunity to examine in greater detail how NAEP inclusion criteria are actually implemented, who gets excluded from NAEP, whether the revisions to the inclusion criteria functioned in the intended ways, and the way in which the determination of "who gets tested in NAEP" will change when accommodations and adaptations are routinely made available.

Moreover, two important technical issues discussed in this report need to be resolved before procedural modifications can be implemented as official NAEP policy. One issue is the validity of results from nonstandard administrations (i.e., administrations in which accommodations were allowed) and comparability to results obtained under standard conditions. Specifically, data obtained under nonstandard conditions may not be able to be summarized and reported according to the same NAEP scale used for results obtained under standard conditions. That is, do scale score results obtained under nonstandard conditions convey the same information about educational achievement as corresponding results obtained under standard conditions?

To date there has been only a modest amount of research on the validity of scores from accommodated assessments. Much of the available research involved higher-achieving student populations taking college admissions and postsecondary tests.[24] Such research findings may not generalize to the primary and secondary school populations, or to assessments such as those encountered in NAEP.

The 1995 NAEP field tests in mathematics and science, though hampered by small sample sizes, did provide some evidence about the comparability of scores from accommodated administrations to those from standard administrations. IRT model fit was examined for LEP students who were administered Spanish and Spanish/bilingual versions of a NAEP test booklet. Such fit was also examined for an aggregate group of students with disabilities who were administered the assessment with some accommodation. Both sets of analyses suggested that results from accommodated sessions may not be well summarized by the same scales used to report NAEP results from the nonaccommodated sessions.[25] Moreover, recent research by Koretz [26] involving the Kentucky Instructional Results Information System provided evidence consistent with that from the NAEP field test. Koretz conducted differential item functioning (DIF) analyses comparing the scores of students with disabilities who received accommodations with the scores of nondisabled students. Statistically significant DIF was found on more than half of the reading and mathematics items studied in grades 4 and 8.

It should be noted that the DIF statistic in the studies by Koretz and as used in chapter 6 of the present report is not employed in the same way as it is in the test development process for NAEP and other examinations. New test items for NAEP are field tested and DIF statistics are calculated between the "focal" group (such as the various ethnic classifications) and the "reference" group (such as the white population). Test questions that show too great a variance might be deleted from the operational examination. In this report and the Koretz work, the DIF statistic is employed as a convenient way to analyze the item by item differences between the accommodated and non-accommodated groups. In the typical course of the test development process, field-tested items

---

[24] Willingham, W.W., Ragosta, M., Bennett, & R.E., Braun, H., Rock, D.A., & Powers, D.E. (1988). *Testing handicapped people.* Boston, MA: Allyn and Bacon.

Wightman, L.F. *Test Takers with Disabilities: A summary from special administrations of the LSAT.* (Research Report No. 93-03). Newtown, PA: Law School Admissions Council.

[25] Anderson, N.E., Jenkins, F.F., & Miller, K.E. (1996). *NAEP inclusion criteria and testing accommodations: Findings from the NAEP 1995 field test in mathematics.* Princeton, NJ: Educational Testing Service.

[26] Koretz, D.M. (1997). *The assessment of students with disabilities.* (CSE Technical Report No. 431). Los Angeles, CA: University of California at Los Angeles, Center for Research on Evaluation, Standards, and Student Testing.

would not have sample sizes sufficient to employ reliably the DIF statistic as a development tool for accommodated vs. non-accommodated groups.

A second issue is the effect of procedural modifications on NAEP's capacity to provide accurate comparisons over time. One of NAEP's goals is to report on trends in academic achievement. Accurately reporting changes requires keeping assessment procedures and instrumentation comparable during the period over which measurement is sought. Modifying inclusion criteria and providing accommodations can significantly expand the number of students with disabilities and LEP students included in NAEP assessments. Although this expansion is desirable, it can cloud the interpretation of changes in achievement over time, since assessments conducted using revised procedures might include results for students who would not have been included in previous assessments.

Organization of This Report

This report provides a follow-up to the analyses presented in the NAEP 1996 report cards in mathematics and science. It contains descriptive results based on NAEP's SD/LEP questionnaire on the background characteristics and educational experiences of students with disabilities and LEP students. It also presents in-depth analyses (on inclusion rates and selected technical characteristics of the assessment) of the effects of the procedural changes designed to increase the participation of special needs students. In particular, data are presented on the effects of revised participation guidelines and the provision of adaptations and accommodations on overall scale score averages and on how well the data for accommodated students fit the NAEP scale. The report does not contain an in-depth examination of the performance on the NAEP assessment of special needs students. The report also does not examine inclusion or performance results for special needs students by state.

The report is organized as follows. Chapter 2 provides descriptive information about the population of students with disabilities in the NAEP school population. Information is given on background characteristics of the students, including the nature and severity of disabilities, aspects of instructional programs, and use of accommodations and adaptations. Chapter 3 presents a detailed examination of the impact that changes in inclusion criteria and provision of accommodations/adaptations had on inclusion rates for students with disabilities. Particular attention is paid to selected subgroups of students and on the actual usage rates of available accommodations and adaptations. The information in chapters 4 and 5 parallels that in chapters 2 and 3, respectively, but focus on LEP students. It should be noted that some students were classified both as students with disabilities and as LEP students. Data from these students were included in the analyses presented in chapters 2 and 3, as well as those presented in chapters 4 and 5.

Chapters 6 and 7 examine some of the key psychometric issues associated with expanding the inclusion of students with disabilities and limited English proficient students in NAEP. Chapter 6 provides information on research methodology, comparisons of item fit statistics, and comparisons of how the overall scaling model fits the data in different samples. Chapter 7 reports on the impact of providing accommodations/adaptations on NAEP estimates of scale score and achievement level distributions for the total population and for selected subgroups of students. Overall summaries of results and conclusions are given in chapter 8.

Description of NAEP 1996 Samples

The descriptive analyses presented in chapters 2 and 4, and the examinations of inclusion rates presented in chapters 3 and 5 are based on the 1996 national mathematics assessment sample. The psychometric analyses presented in this report are based on both the 1996 national NAEP mathematics and science samples. Because of a desire to keep the scope of chapters 6 and 7 within reasonable limits, data from the 1996 state assessments are not included. The mathematics and science assessment results presented in this report are based on nationally representative probability samples of fourth-, eighth-, and twelfth-grade students. As discussed above, there were three distinct mathematics samples and two distinct science samples. Table 1.2 presents the mathematics sample sizes by grade for each of the sampling conditions (S1, S2, and S3) and for the students with disabilities and well as those with limited English proficiency.

In grade 4, nearly 1,200 students with disabilities were sampled. Students with limited English proficiency totaled 754. In grade 8, nearly 1,400 students with disabilities were sampled, while students with limited English proficiency totaled 650. At grade 12, sample sizes for special needs students were smaller, with about 1,000 students with disabilities sampled and just over 500 students with limited English proficiency sampled.

Table 1.2 — NAEP 1996 national student sample sizes by sample type: Mathematics

	Students with disabilities				Students with limited English proficiency			
	All samples	S1[1]	S2[2]	S3[3]	All samples	S1	S2	S3
Grade 4								
Assessed standard	567	206	189	172	314	75	125	114
Assessed with accommodations	143	—	—	143	108	—	—	108
Excluded	484	153	222	109	332	67	179	86
Total	1,194	359	411	424	754	142	304	308
Grade 8								
Assessed standard	675	161	287	227	383	68	182	133
Assessed with accommodations	147	—	—	147	42	—	—	42
Excluded	569	149	237	183	225	38	136	51
Total	1,391	310	524	557	650	106	318	226
Grade 12								
Assessed standard	379	103	169	107	367	38	151	178
Assessed with accommodations	73	—	—	73	12	—	—	12
Excluded	556	108	242	206	129	9	82	38
Total	1,008	211	411	386	508	47	233	228

[1] 1900-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

NOTE: — Not applicable because accommodations were not offered.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table 1.3 presents the sample sizes for the special needs students in the 1996 science assessment. Note that while the totals are similar to those for mathematics in table 1.2, the S1 condition did not apply because, as noted earlier, the 1996 science assessment was based on a new framework, which was different from the framework guiding earlier science assessments. The absence of the S1 condition, however, allowed for larger numbers of students in the S2 condition in science than in mathematics. These larger samples in S2 provided a fortuitous opportunity in the psychometric work (see chapters 6 and 7) to do randomized split-half analyses, which served as a kind of control against which to compare the results of the S2 versus S3 group comparisons.

Table 1.3 — NAEP 1996 national student sample sizes by sample type: Science

THE NATION'S REPORT CARD naep

	Students with disabilities				Students with limited English proficiency			
	All samples	S1[1]	S2[2]	S3[3]	All samples	S1	S2	S3
Grade 4								
Assessed standard	540	—	348	192	399	—	261	138
Assessed with accommodations	159	—	—	159	36	—	—	36
Excluded	570	—	425	145	589	—	393	196
Total	1,269	—	773	496	1,024	—	654	370
Grade 8								
Assessed standard	672	—	449	223	364	—	217	147
Assessed with accommodations	137	—	—	137	35	—	—	35
Excluded	558	—	314	244	257	—	156	101
Total	1,367	—	763	604	656	—	373	283
Grade 12								
Assessed standard	320	—	201	119	345	—	174	171
Assessed with accommodations	73	—	—	73	5	—	—	5
Excluded	524	—	321	203	188	—	136	52
Total	917	—	522	395	538	—	310	228

[1] 1900-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

NOTE: — Not applicable because accommodations were not offered.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table 1.4 displays the numbers of schools that provided data for both the mathematics and science analyses in this report. Approximately 300 schools at each grade level provided data in mathematics. Somewhat fewer schools (242-290) provided data for the science assessments of special needs students.

Table 1.4 — NAEP 1996 mathematics and science: Number of schools by sample type providing data on special needs students

	Mathematics				Science			
	All samples	S1[1]	S2[2]	S3[3]	All samples	S1	S2	S3
Grade 4	313	80	117	116	290	—	161	129
Grade 8	294	73	112	109	242	—	126	116
Grade 12	302	56	125	121	280	—	162	118

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

NOTE: — Not applicable because accommodations were not offered.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics and Science Assessments.

Sample Weights, Sampling Errors and Nonresponse Consideration

NAEP's data collection and sampling contractor (Westat) used a complex multistage design to select the NAEP samples.[27] Each selected student who participated in the assessment represents a portion of the population of interest. Sampling weights must be used in analysis to allow for a proper comparison of results across sample types and to make valid inferences from the student samples to the respective populations from which they were drawn. Unless otherwise noted, all analyses presented in this report were conducted using appropriate sampling weights provided by the NAEP data collection and Westat. Details of the computations of weights may be found in the forthcoming *NAEP 1996 Technical Report*.[28]

[27] Allen, N.L., Carlson, J.E., & Zelenak, C.A. (forthcoming). *The NAEP 1996 technical report.* Washington, DC: National Center for Education Statistics.

[28] Ibid.

The 1996 design required that different sets of weights be created for different purposes. The mathematics assessment was a trend assessment. For the purposes of comparing the 1996 results with earlier mathematics assessments, a special set of “reporting weights” was required. The mathematics reporting sample was made up of all of the students from S1 and S2 who were not classified as either SD or LEP students, plus the SD and LEP students from S1 who were determined to be capable of taking the assessment under standard administration conditions. No accommodations were provided to students in these two samples. The resulting reporting population is, by design, equivalent to that used for the 1992 mathematics assessment. The 1996 science assessment was based on a new framework. There was therefore no need to administer that assessment under conditions similar to those in previous assessment years. A decision was made to establish the S2 sample as the official reporting sample for the 1996 science results. The analyses presented in the 1996 mathematics and science report cards are based on these reporting samples. Selected results in chapters 6 and 7 also make use of the reporting samples and their respective weights.

For analyzing data within S1, S2, and S3, Westat provided a set of “modular weights.” These weights, when used in the analyses reported here, result in statistics that provide appropriate estimates of the population parameters of interest. That is, the sample results are estimates of what might be expected if all members of the population were assessed under the conditions of the sample type in question. Most of the results presented in chapters 3 and 5 and all the results presented in chapters 6 and 7 are based on these modular weights.

For the descriptive analyses presented in chapters 2 and 4 and a small number of the analyses presented in chapters 3 and 5, data from the three mathematics samples were pooled to provide a larger combined data set. A special set of combined weights was produced by ETS. The combined weights were derived from the modular weights and were designed to provide unbiased estimates of population characteristics for all students with disabilities and LEP students (i.e., students assessed as well as excluded). The combined weights were derived as follows:

- Let m(ij) refer to the modular weight associated with the ith individual from sample j. So, for example, m(11) is the modular weight for person 1 from the S1 sample.
- Let n(1), n(2), and n(3) refer to the student sample sizes associated with S1, S2, and S3, respectively. Further, define N as the total combined unweighted sample size (i.e., N = n(1) + n(2) + n(3)).
- Combined weights were defined as: c(ij) = m(ij) [n(j)/N]

Essentially, the combined weights represent a rescaling of the modular weights, with the scaling factor defined as the ratio of the student sample size within sample type to the total student sample size (i.e., the sample size across all three sample types).

The results presented in this report are estimates of group and subgroup characteristics and performances based on samples. Since they are estimates, their associated degree of uncertainty should be taken into account. Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate. NAEP uses a jackknife replication procedure to estimate standard errors. Jackknife estimates of the standard errors for the results reported in chapters 2 through 5 are presented in appendix G. When the percentages or average scale scores of certain groups are compared, the observed differences (or lack thereof) should not be relied on solely. The standard errors of the statistics, as well as their associated degrees of freedom, should be taken into account. Therefore, the comparisons discussed in this report are based on statistical tests that consider the magnitude of the differences, their estimated standard errors, and the degrees of freedom associated with the estimates. All statistical tests were two-tailed at the .05 level of significance. In chapter 7, where large numbers of comparisons were conducted, multiple-comparison procedures were employed to control the family wide error rate at .05. Further detail on these multiple comparison procedures is provided in that chapter.

Along with the collection of assessment data, staff members at participating schools who were knowledgeable about the special needs students were asked to fill out a detailed questionnaire (see appendix B) for each special needs student. Across the mathematics and science assessments, the missing (nonreturned) questionnaire rates for students with disabilities ranged from 12 to 22 percent. The nonreturn rates for students with limited English proficiency were somewhat higher, ranging from 23 to 32 percent. The extent of missing questionnaire data raised concerns about the degree to which the results presented in chapters 2 through 5 are, in fact, representative of the full NAEP special needs population. Therefore, a series of analyses was conducted using the national mathematics samples to compare the subset of special needs students with returned questionnaires to all special needs students with returned questionnaires. The two groups were compared with respect to several school-level and student-level demographic variables. These analyses (see appendix A) indicate relatively minor differences between the full group of special needs students and those with returned questionnaires. Based on these analyses, it seems reasonable to assume that the results are representative of the population of fourth-, eighth-, and twelfth-grade special needs students attending schools in the NAEP sampling frame. However, it is possible that the results based on the returned questionnaires may reflect other kinds of response biases (e.g., attitudes about students with disabilities; lack of knowledge about these students) that the NAEP program was not able to measure.

Chapter 2

Students with Disabilities: A Description of the NAEP Population

This chapter presents a description of the nation's fourth-, eighth-, and twelfth-grade students with disabilities attending NAEP-participating schools, based on SD/LEP questionnaire results obtained as part of NAEP's 1996 assessment.[1] The results presented in this chapter were obtained by pooling the SD/LEP questionnaire data from the S1, S2, and S3 school samples (see pages 8-9) of the national mathematics samples, with appropriate adjustments to the sampling weights (see tables 1.2 and 1.4 for student and school sample sizes, respectively), to provide a single set of "best" estimates for describing students with disabilities in NAEP schools. Results for the mathematics sample are presented in lieu of those for the science assessment because school and student sample sizes for the combined mathematics sample were slightly larger than those for the science sample. Because the 1996 science assessment was conducted in the same schools as the mathematics assessment, demographic results based on those samples differ little from those presented here.

---

[1] NAEP's sampling frame does not include ungraded schools, public/private day schools exclusivly serving students with disabilities, or public/ private residential facilities.

Table 2.1 — Percentage distribution of job titles of students with disabilities questionnaire respondents by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Type of respondent	Grade 4	Grade 8	Grade 12
Principal/assistant principal	1	2	2
Special education teacher	50	66	64
Bilingual education/ESL teacher	0	0	0
Classroom teacher	30	7	4
Other	3	11	11
Blank	19	14	18

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

As noted in chapter 1, return rates for the SD/LEP questionnaire (see page 11) based on the 1996 national mathematics samples ranged from 82 percent (at grade 4) to 78 percent (at grade 12), and comparable return rates were observed for the national science samples. Analyses, presented in appendix A, indicate relatively minor differences between the full group of students with disabilities and those with returned questionnaires with respect to a number of school- and student-level variables. Based on these analyses it is reasonable to assume that the results presented in this and the ensuing chapters are representative of the population of fourth, eighth, and twelfth graders with disabilities attending schools in the NAEP sampling frame.

A knowledgeable school staff member (e.g., a member of the student's Individualized Educational Plan (IEP) team) was asked to fill out the SD/LEP questionnaire for each student with a disability, regardless of whether the student was judged capable of participating in the NAEP assessment. Table 2.1 presents information on the position of the person(s) who completed the questionnaires.

For 19 percent of grade-four students, the position of the person filling out the questionnaire was not indicated. Almost all of the remaining 81 percent were filled out by the special education teacher (50 percent) or the student's classroom teacher (30 percent). The position of the person filling out the questionnaire was also not indicated for 14 percent of the grade-eight students and 18 percent of the grade-twelve students. For both of these grades, the special education teacher typically filled out the questionnaire. However, for 11 percent of the students at each of these grades, the respondent was a person with a position not among those listed on the questionnaire. Anecdotal reports from the field suggest that this was typically a classroom or special education aide working closely with the student or a school counselor (e.g., the school psychologist).

Based on estimates from the NAEP 1996 mathematics assessment, 360,000 (11 percent) of fourth graders, 277,000 (9 percent) of eighth graders, and 130,000 (5 percent) of twelfth graders attending schools in the NAEP sampling frame were identified as having one or more disabilities. At grades 4 and 8, 97 percent of these students had IEPs, while 3 percent had equivalent classifications. The corresponding percentages at grade 12 were 95 and 5 percent, respectively. What kinds of disabilities did they have and how severe were these disabilities? Obtaining answers to such questions is less than straightforward. As discussed in McDonnell *et al.*, no official special education classification is used uniformly across the nation.[2] Definitions of disability categories and classification criteria vary across states, as well as across districts within states. Federal regulations define thirteen disability categories and individual students receiving special services under IDEA are required to indicate one of these as the primary disability category. However, some students no doubt have multiple disabilities.

The 1996 NAEP assessment collected some information on the nature and severity of student disabilities. For each student with a disability in the NAEP sample, questionnaire recipients were presented a list of disabilities and asked to select all that applied. Table 2.2 presents the percentages of students with disabilities identified for each of the categories. Percentages are given for all students with disabilities, males, and females. It should be noted that the percentages collected in the NAEP survey are not directly comparable to those collected under federal regulations. For example, the list of disabilities presented in the NAEP questionnaire is similar but not identical to the thirteen federal reporting categories. In addition, federal statistics are collected on each student's primary disability classification and are disaggregated by age. The NAEP samples were defined on the basis of grade, and do not include students in ungraded special centers. Further, NAEP questionnaire recipients were asked to indicate all disabilities that applied. More than one disability was indicated for 29 percent of fourth graders, 19 percent of eighth graders, and 18 percent of twelfth graders with disabilities.

---

[2] McDonnell, L.M., McLaughlin, M.J., & Morison, P. (Eds.). (1997). *Educating one & all: Students with disabilities and standards based reform.* Washington, DC: National Academy of Science, National Research Council.

Table 2.2 — Percentage of students with disabilities by selected disability type by grade and gender: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Which of the following describes this student's disability?	Grade 4			Grade 8			Grade 12		
	All students	Male	Female	All students	Male	Female	All students	Male	Female
Learning disability	72	74	67	76	77	73	72	74	68
Speech/language impairment	26	24	31	8	6	10	7	6	8
Mental or cognitive impairment	17	14	24	18	14	24	25	22	31
Emotional disturbance	9	11	6	12	13	11	9	9	8
Hard of hearing	1	1	1	1	0	1	2	1	4
Deaf	0	0	1	0	0	0	1	1	1
Visual impairment/blindness	2	2	2	1	0	2	2	1	4
Orthopedic impairment	2	2	2	2	1	3	2	2	3
Autism	1	0	2	0	1	0	0	0	0
Traumatic brain injury	0	0	0	0	0	0	1	1	1
Other	6	6	6	5	5	6	4	4	4

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Despite the above-mentioned methodological differences, there are a number of similarities between NAEP results and other federal statistics. At all three grades, the vast majority of students with disabilities was classified into at least one of four of the listed categories (learning disability, speech/language impairment, mental or cognitive impairment, and emotional disturbance). Federal statistics suggest that, in the 1995-1996 school year, over 90 percent of children with disabilities in school settings had primary disabilities in one of these four categories.[3] Learning disability was by far the most frequently reported category. At all three grades, close to three of four students with disabilities were identified as having a learning disability. Federal statistics indicate a somewhat lower incidence (50 percent) had a learning disability as their primary disability. Differences in the NAEP results may be due, among other things, to allowing multiple categories to be indicated for each student. Speech/language impairments were fairly prevalent among fourth graders with disabilities (26 percent), but appear less common at the higher grades (8 and 7 percent at grades eight and twelve, respectively).

---

[3] U.S. Department of Education. (1997). *Nineteenth annual report to Congress on the implementation on the Individuals with Disabilities Education Act.* Washingtion, DC: Author.

As shown in table A4.a (in appendix A), males are overrepresented among students with disabilities, making up about two-thirds of the group. However, as shown in table 2.2, the pattern of female disabilities is similar to that of males. Learning disability was the most commonly reported disability for both groups.

Students within a disability category may differ substantially with respect to the size of any intellectual, communicative, or behavioral deficit, the number of areas such deficits are evidenced in, and the complexity of educational interventions required. All three of these factors no doubt contribute to judgments about the severity of student disabilities. Recipients of NAEP's SD/LEP questionnaire were asked to indicate the degree of each student's disabilities with respect to four possible categories: mild, moderate, severe or profound. Results for this question are presented in table 2.3 for all students with disabilities, students with cognitive or mental impairments, and students with disabilities other than cognitive or mental impairments.

About half of the students with disabilities at each grade (46 percent at grade four, 53 percent at grade eight, and 49 percent at grade twelve, respectively) were described as having mild disabilities. The remaining half at each grade were almost all categorized as having moderate to severe disabilities. Very few students receiving educational services at NAEP schools (1 percent at grade four, 1 percent at eight, and 3 percent at grade twelve) were judged to have profound disabilities. It also appears from table 2.3 that severe or profound disabilities were more common among students with cognitive or mental impairments than among students with other disabilities. Given that students with disabilities who were eligible for the NAEP sample received at least some of their educational services in schools with nondisabled students, it is probably not surprising to find that nearly 85 percent of these students with disabilities at each grade were classified as having mild or moderate disabilities. Those with more severe disabilities, and hence in need of extensive specialized care, were more likely to attend separate day schools, residential facilities, or hospital or homebound programs that are not included in NAEP samples. For those within the NAEP sample, the likelihood of exclusion from assessment increases with the reported severity of disability. This is examined more fully in the next chapter (Table 3.6).

Table 2.3 — Percentage distribution of degrees of disabilities by type of disability and grade: NAEP 1996 mathematics sample

What is the degree of this student's disability?	Grade 4			Grade 8			Grade 12		
	All students with disabilities	Students with mental impairment	Students with other disabilities	All students with disabilities	Students with mental impairment	Students with other disabilities	All students with disabilities	Students with mental impairment	Students with other disabilities
Profound	1	3	0	1	4	1	3	7	2
Severe	14	23	12	12	24	9	11	20	8
Moderate	39	40	39	34	38	33	36	41	35
Mild	46	34	49	53	35	57	49	32	55

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

As noted earlier, the total percentage of students with disabilities differs across the three grades. The percentage of grade 12 students with disabilities is less than half of the corresponding percentage at grade 4. Differences in these percentages are no doubt caused by a number of factors. First, the NAEP samples represent different cohorts of students, each separated from the other by four years. Since diagnoses of significant and physical disabilities are often made prior to the onset of schooling or in the early years of formal education, the larger percentages at grade 4 could reflect a greater current tendency to make such diagnoses or greater ability to recognize such deficits. Furthermore, some students with mild disabilities may have had them remediated in the earlier grades, and hence, lost the disability classification in the later grades. Lastly, students with more severe disabilities may have dropped out of school or been relocated to special schools for the disabled that are not included in the NAEP samples. In light of the factors above, it is somewhat surprising that the results in tables 2.2 and 2.3 suggest that, with one or two exceptions, the nature and level of severity of student disabilities exhibited similar patterns in all three grades.

Educational Experiences and Levels of Performance

Students with disabilities experience a wide range of educational programs. As noted in McDonnell *et al.* (1997),

> Having a disability, mild or severe, can affect a child's schooling in many ways. It can affect where children are educated, whether they have the same goals for schooling as students without disabilities, and whether they participate in all of the general education curriculum, some of it, or none of it. Furthermore, it can influence whether they can be taught by the same methods and with the same tools and equipment as other students, and whether they can be evaluated in the same ways.[4]

There is an acknowledged paucity of representative data on the range and degree of participation in various aspects of the general education curriculum among students with disabilities. However, the results from NAEP's SD/LEP questionnaire provide some representative data on the school experiences of these students.

Federal statistics for the 1995-96 school year suggest that all but 5 percent of the nation's students with disabilities attend regular schools.[5] Students with disabilities attending regular schools receive educational services in a variety of ways. Some (44 percent) receive almost all their instruction in regular classrooms with their nondisabled peers, others (22 percent) spend most of their time in separate special education classrooms, and still others (29 percent) split their time more evenly between regular classrooms and resource rooms. SD/LEP questionnaire recipients were asked to indicate the percentage of time each student with a disability spends mainstreamed in academic subjects. Results for this question are shown in table 2.4. Results are shown separately for all students and for the four most frequent disability types: learning disability, cognitive or mental impairment, speech/language impairment, and emotional disturbance.

Regardless of grade level, about half (i.e., 47 to 54 percent) of all students with disabilities were mainstreamed in academic subjects at least 80 percent of the time. At grade 4, mainstreaming in academic subjects appeared to be less prevalent among students with mental or cognitive disabilities than it was among students in the other major disability categories. At grades 8 and 12, students with learning disabilities appeared more likely to be mainstreamed at least 80 percent of the time than were students in the other disability categories, and students with mental or cognitive impairments or speech/language impairments appeared least likely to be mainstreamed. At the other extreme, the use of self-contained classrooms for academic instruction appeared to be more prevalent at grades 8 and 12 than at grade 4. Among students with disabilities, 31 percent of grade twelve students and 27 percent of grade eight students were mainstreamed in academic subjects less than 40 percent of the time. This contrasts with 19 percent of grade four students with disabilities. At all three grades, students with learning disabilities appeared less likely to be mainstreamed less than 40 percent of the time than were students in the other major disability categories.

---

[4] McDonnell, L.M., McLaughlin, M.J., & Morison, P. (Eds.). (1997). *Educating one & all: Students with disabilities and standards based reform.* Washington, DC: National Academy of Science, National Research Council.

[5] U.S. Department of Education. (1997). *Nineteenth annual report to Congress on the implementation on the Individuals with Disabilities Education Act.* Washington, DC: Author.

Table 2.4 — Percentage of time students with disabilities are mainstreamed in academic subjects by type of disability and grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

What percentage of time is this student mainstreamed (i.e., with his/her nondisabled peers) in academic subjects (e.g., mathematics, reading/language arts, science)?	Less than 40%	40% to 79%	At least 80%
Grade 4			
All Students	19	26	54
Students with a:			
Learning disability	16	29	55
Cognitive impairment	41	35	25
Language impairment	24	22	54
Emotional disturbance	28	25	47
Grade 8			
All Students	27	26	47
Students with a:			
Learning disability	23	28	50
Cognitive impairment	64	22	14
Language impairment	45	29	26
Emotional disturbance	36	30	35
Grade 12			
All Students	31	20	49
Students with a:			
Learning disability	21	22	56
Cognitive impairment	70	19	10
Language impairment	68	18	15
Emotional disturbance	37	20	43

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Grade Levels of Curriculum Experienced. Another important aspect of schooling on which students with disabilities differed is the degree to which they are exposed to the same general curriculum as students without disabilities. There are at least two aspects to this issue. One aspect involves the grade level of material in which students are receiving instruction. Specifically, are students with disabilities more apt to be receiving instruction that is below grade level? A second aspect, independent of grade level, involves the nature of the instructional content. Do students with disabilities receive the same instructional content as nondisabled students receiving instruction at the same grade level?

To provide information on level of instruction, SD/LEP questionnaire recipients were asked to indicate the grade level of instruction each student was receiving in three distinct areas: reading/language arts, mathematics, and science. Results for these questions are given in table 2.5. Despite some subject-area-specific peculiarities, there are a number of similarities in results across the subject areas.

In reading/language arts, fewer than half of all students with disabilities received instruction at grade level. Percentages of students with disabilities who received grade-level instruction ranged from 48 percent at grade eight to 37 percent at grade four. In mathematics and science, the situation appears to be slightly better at the two lower grades. About half of the grade four and grade eight students with disabilities received grade-level instruction. About seventy percent of students with disabilities received grade-level instruction in science at grades four and eight. The situation at grade 12 in mathematics and science was not so encouraging. Only 26 percent of grade-twelve students received grade-level instruction in mathematics and a similar percentage (22 percent) received grade-level instruction in science.

Moreover, in all three subject areas, students with disabilities at the higher grades appear to be further behind with respect to instructional level. In mathematics, 16 percent of grade-four students with disabilities received instruction that is two or more years below grade level. For grade-eight students, the corresponding percentage more than doubles to 35 percent. Among grade-twelve students with disabilities, 50 percent received instruction that is two or more years below grade level. A similar pattern of results is evident in science instruction for students with disabilities. The situation in reading/language arts differs only somewhat from that observed in mathematics and science. At grades four and eight, about 40 percent of students with disabilities received reading/language arts instruction that is two or more years below grade level. However, at grade twelve this group increased to 52 percent.

It should be noted that educational significance of the results in Table 2.5 could be more readily interpreted if comparable data were available for students without disabilities. At least one study[6] found that when elementary school teachers were asked to provide grade level estimates for students without disabilities, spreads of 4 or more grades were reported. Unfortunately, NAEP did not collect comparable data for students without disabilities.

---

[6] Thurlow, M.L., Christenson, S.L., Ysseldyke, J.E., Franklin, M.J. , & Shriner, J.G. (1989). *Student academic responses under varying group size and composition, skill range, and student teacher ratios in general education classrooms.* (Research Report No. 25). Minneapolis, MN: University of Minnesota, Instructional Alternatives Project.

Table 2.5 — Percentage distribution of grade level of instruction in reading/language arts, mathematics, and science for students with disabilities by grade: NAEP 1996 mathematics sample

What grade level of instruction is the student currently receiving in:	Grade 4	Grade 8	Grade 12
Reading/language arts			
At/above grade level	37	48	38
One year below grade level	22	7	2
Two or more years below grade level	38	40	52
Missing	3	5	7
Mathematics			
At/above grade level	55	52	26
One year below grade level	24	8	3
Two or more years below grade level	16	35	50
Missing	4	5	20
Science			
At/above grade level	70	68	22
One year below grade level	8	3	3
Two or more years below grade level	13	22	38
Missing	8	7	37

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Similarity of Content. To provide information on the similarity of instructional content for students with disabilities and for nondisabled students, SD/LEP questionnaire recipients were asked to indicate by subject area (i.e., reading/language arts, mathematics, and science) whether each student was receiving the same curriculum content as their nondisabled peers at the same grade level. Results for these questions are given in table 2.6 for all students with disabilities, for students with disabilities receiving instruction at or above grade level, and for students with disabilities receiving instruction below grade level.

In all three subject areas and at all three grades, the majority of students with disabilities received the same content of instruction as their nondisabled peers at the same grade level. At grades four and eight in science, close to 85 percent of students with disabilities received the same curriculum content as nondisabled students. When coupled with the fact that nearly three-quarters of fourth- and eighth-grade students with disabilities received instruction in science that is at grade level, the results in table 2.6 suggest a good deal of comparability in the science curriculum for both regular and disabled students. However, for all three subjects the percentages appear smaller at grade twelve, ranging from 55 percent in mathematics to 62 percent in science.

The results in table 2.6 also suggest that the situation differs dramatically for students receiving grade-level instruction compared to students receiving instruction that is below grade level. In all three subject areas, almost all students who received instruction that is at or above grade level received the same curriculum content as their nondisabled peers. In contrast, with one exception, fewer than half of those students with disabilities who received below grade-level instruction were taught the same curriculum content as their nondisabled peers. This pattern of results is similar across all three grade levels.

Table 2.6 — Percentage of students with disabilities receiving the same curriculum content as nondisabled students at the same grade level, by grade level of instruction and grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Instructional areas	Grade 4	Grade 8	Grade 12
Reading/language arts			
All Students	62	67	59
Students receiving instruction:			
At/above grade level	95	95	95
Below grade level	41	38	33
Mathematics			
All Students	75	68	55
Students receiving instruction:			
At/above grade level	98	96	95
Below grade level	45	33	34
Science			
All Students	86	84	62
Students receiving instruction:			
At/above grade level	98	98	98
Below grade level	52	48	42

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Instruction Areas in Special Education Programs. The vast majority of students with disabilities have IEPs which, among other things, delineate the academic, behavioral, vocational, and social skills domains in which they will receive special education services. SD/LEP questionnaire recipients were presented a list of seven areas, and an "other" option, and asked to indicate in which of these areas each student received instruction as part of their special education program. At least one area of special education instruction was indicated for 94 percent of the fourth-grade students with disabilities, 87 percent of eighth-grade students with disabilities, and 80 percent of twelfth-grade students with disabilities (data not shown). The percentages for each instructional area, including "other," are presented in table 2.7.

At all three grades, the most common areas of special education instruction were in language development, reading, and mathematics. The percentages of students receiving special instruction in language development and reading appear higher at grade four than at the other grades. A more variable pattern exists for mathematics. It is interesting to note, however, that in all three of these instructional areas, about 40 percent of all grade-twelve students with disabilities are receiving services.

With one or two exceptions, special instruction in the other areas appeared to be less common among students with disabilities. Special instruction in speech appeared fairly common (27 percent) among grade-four students with disabilities, but less so among grade-eight (10 percent) and grade-twelve students (6 percent). This pattern of results mirrors the earlier-noted finding that speech/language impairments were found more frequently among fourth graders than among students at the higher grades. As might be expected, vocational education was almost nonexistent at the fourth grade, but appeared more common at the higher

Table 2.7 — Percentage of students with disabilities receiving instruction in selected areas as part of their special education programs by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD

Instructional areas	Grade 4	Grade 8	Grade 12
Language development	56	47	39
Reading	73	57	41
Mathematics	49	58	39
Speech	27	10	6
Self-control and deportment	16	14	11
Personal care and basic life skills	10	11	14
Vocation education	2	8	38
Other	10	19	20

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

grades. Among the grade-twelve students with disabilities, 38 percent received some form of vocational education as part of their special education programs.

Levels of Achievement. There is little in the way of solid, nationally representative data indicating levels of achievement among students with disabilities. The data that do exist suggest that students with disabilities exhibit lower levels of academic achievement, high school graduation, postsecondary school enrollment, and employment.[7] NAEP reading results from 1992 and 1994, NAEP mathematics results from 1990, 1992, and 1996, and NAEP science results from 1996 all show that the subset of students with disabilities included in NAEP reporting samples (i.e., those students who participated without accommodations or adaptations) have substantially lower test scores than their nondisabled peers.[8]

As part of the 1996 NAEP, SD/LEP questionnaire recipients were asked to indicate for each student the observed grade level of performance in three academic areas: reading/language arts, mathematics, and science. Results for these questions are reported in table 2.8. At all three grades and in all three subject areas, some students with disabilities are reported to be achieving at or above grade level. However, the NAEP questionnaire results are consistent with the existing research in suggesting low levels of academic achievement for the majority of these students.

In all three grades, about 75 percent of students with disabilities were judged to be performing below grade level in reading/language arts. The percentage of students two or more years below grade level appeared to increase from 44 percent at grade four to 63 percent at grade twelve. In mathematics, grade-four performance levels appear to be somewhat higher than those for reading/language, with 39 percent of students with disabilities reported to be performing at or above grade level. However, reports on mathematics performance at the higher grades follow a pattern similar to that shown in reading, with increasing percentages of students reported to be performing two or more years below grade level as age increases. Among grade-twelve students with disabilities, 54 percent were reported to be performing two or more years below grade level in mathematics, compared with 24 percent at grade four.

Questionnaire recipients appeared to be less knowledgeable about student performance levels in science, particularly at grades eight and twelve, where such levels were not known for 19 and 38 percent of the students, respectively. This lack of knowledge at grade 12 may reflect the fact that at least some of these students are less likely to be taking science at grade 12. For those students for whom science performance levels were reported, a pattern of results emerges that is similar to that in mathematics. Performance levels in science are estimated to be somewhat higher than those in reading/language arts at grade 4. Of grade-four students with disabilities, 46 percent were reported to be performing at or above grade level. However, at the higher grades, the incidence of report of below-grade-level performance appears to be higher. At grade twelve, 39 percent of the students with disabilities were reported to be performing two or more years below grade level in science.

---

7 Smith, T.M., Young, B.A., Choy, S.P., & Alsalem, N. (1997). *The condition of education 1997*. (NCES Publication No. 97-388). Washington, DC: National Center for Education Statistics.

8 Results for each of the subject areas are available on the World Wide Web at the NCES website (*http://nces.ed.gov/nationsreportcard)* in the Student Data sections of the Summary Data Tables.

Table 2.8 — Percentage distribution of estimated grade level of performance by students with disabilities in reading/language arts, mathematics, and science by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

At what grade level is this student currently performing in :	Grade 4	Grade 8	Grade 12
Reading/language arts			
Above grade level	2	1	1
At grade level	19	21	18
One year below grade level	31	18	9
Two or more years below grade level	44	51	63
I don't know.	4	9	9
Mathematics			
Above grade level	1	1	3
At grade level	38	30	15
One year below grade level	32	18	8
Two or more years below grade level	24	43	54
I don't know.	6	7	20
Science			
Above grade level	0	1	2
At grade level	46	36	16
One year below grade level	20	15	6
Two or more years below grade level	19	30	39
I don't know.	14	19	38

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

It is interesting to compare the results on estimated performance levels in table 2.8 to the report on instructional levels given in table 2.5. Across all grades and subjects, reported performance levels appear to lag behind reported instructional levels. For example, 70 percent of students with disabilities were reported to be receiving instruction at or above grade level. However, only 40 percent were estimated to be performing at or above level. As noted earlier, comparable data for students without disabilities would help assess the educational importance of these results. However, such data were not collected.

Accommodations and Adaptations Used by Students with Disabilities. As noted in McDonnell *et al*. (1997)[9], "One of the avenues for increasing participation of students with disabilities in assessments is allowing accommodations." According to Bond (1996)[10] as of 1995 there were 35 states that permitted the use of special testing conditions and accommodations in their state testing programs for students with disabilities.

Thurlow[11] characterized accommodations currently in use as falling into four broad categories: (1) ***presentation accommodations*** (e.g., Braille forms, aural presentations), (2) ***response accommodations*** (e.g., sign language, oral responses), (3) ***setting accommodations*** (e.g., small-group testing and individual testing), and (4) ***timing accommodations*** (e.g., extended time). SD/LEP questionnaire recipients were presented a list of each of these types of accommodations and asked to indicate which ones were generally used in achievement testing for each student with a disability. Respondents were permitted to check all that applied because experience with NAEP and other testing programs indicates that students typically receive combinations of these accommodations.

Table 2.9 displays the percentages of students with disabilities who generally used an accommodation or adaptation in achievement testing and the percentages who used one or more of each of the accommodation types. Results are similar across grades both with respect to the percentage of accommodations used and to the percentage of each accommodation type used. Across the three grades, respondents reported that 42 to 44 percent of students with disabilities received some form of accommodation or adaptation in testing.

[9] McDonnell, L.M., McLaughin, M.J., & Morison, P. (Eds). (1997). *Educating one & all: Students with disabilities and standard based reform.* Washington, DC: National Academy of Science, National Research Council.

[10] Bond, L.A., Braskamp, D., & Roeber, E.D. (1996). *The status of state assessment programs in the United States*. Oakbrook, IL: Council of Chief State School Officers, North Central Regional Educational Laboratory.

[11] Thurlow, M.L., Ysseldyke, J.E., & Silverstein, B. (1993). *Testing accommodations for students with disabilities: A review of the literature*. (Synthesis Report No. 4). Minneapolis, MN: University of Minnesota, National Center for Educational Outcomes.

Table 2.9 — Percentage of students with disabilities using one or more accommodations for achievement testing by grade: NAEP 1996 mathematics sample

Percentage of students with disabilities receiving:	Grade 4	Grade 8	Grade 12
Any accommodation	43	42	44
Presentation accommodation	37	33	34
Timing accommodation	35	37	41
Setting accommodation	38	36	35
Response accommodation	18	15	19

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Between 33 and 41 percent of students with disabilities used presentation, timing, or setting accommodations. Response accommodations appear to be used less frequently, ranging from 15 percent at grade eight to 19 percent at grade twelve.

Table 2.10 presents a list of presentation accommodations and the percentages of students for which each was used. The most frequently reported presentation accommodations among students with disabilities were reading directions aloud, reading problems aloud, and providing assistance with directions. The remaining presentation accommodations on the list were used quite infrequently, although 4 to 5 percent of students apparently used some accommodation that was not included in the questionnaire choices.

Table 2.10 — Percentage of students with disabilities receiving selected presentation accommodations and adaptations in achievement testing by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Type of presentation accommodation:	Grade 4	Grade 8	Grade 12
Read directions aloud	33	26	25
Read problems aloud	26	22	20
Assistance with directions	18	17	19
Use of taped version of test	2	1	4
Accommodation not listed on the questionnaire*	4	4	5

* At all three grades, less than half of 1 percent indicated each of the following accommodations: signing of directions; Braille edition of test; large-print edition of test; use of magnifying equipment.

NOTE: Questionnaire respondents were instructed to choose all responses that apply.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table 2.11 presents a list of timing accommodations and the percentages of students for which each was used. Extended testing time was the most commonly reported timing accommodation. Across the grades, extended time was reported being used for 32 to 39 percent of students with disabilities. Other timing accommodations, such as allowing additional breaks during testing and allowing testing sessions to extend over several days, were reported less frequently.

Table 2.11 — Percentage of students with disabilities receiving selected types of timing accommodations in achievement testing by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD NAEP

Type of timing accommodation	Grade 4	Grade 8	Grade 12
Extended time	32	34	39
More breaks during test	11	9	9
Test sessions over several days	7	6	9
Accommodation not listed on the questionnaire	2	1	0

NOTE: Questionnaire respondents were instructed to choose all responses that apply.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table 2.12 presents a list of setting accommodations and the percentages of students for which each was used. The most typical setting accommodation was testing students in small groups. Among students with disabilities, 30 percent of fourth graders, 27 percent of eighth graders, and 26 percent of twelfth graders were tested in small groups.

Table 2.12 — Percentage of students with disabilities receiving selected setting accommodations in achievement testing by grade: NAEP 1996 mathematics assessment

THE NATION'S REPORT CARD NAEP

Type of setting accommodation	Grade 4	Grade 8	Grade 12
Test in small group	30	27	26
Test individually	13	10	16
Accommodation not listed on the questionnaire	3	4	2

NOTE: Questionnaire respondents were instructed to choose all responses that apply.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table 2.13 presents a list of response accommodations and the percentages of students for which each was used. In general, none of the listed response accommodations was used with great frequency among students with disabilities. The use of a Braille booklet or talking calculator, though infrequent at grade 4, appeared to be a more common response accommodation at grades 8 and 12. Similarly, the use of a computer to respond was reported only 2 percent of the time in grades four and eight, but 5 percent of the time in grade twelve. It should be noted that other types of response accommodations not on the NAEP list were in use. Among students with disabilities, it was reported that 4 percent of grade-four students, 2 percent of grade-eight students, and 5 percent of grade-twelve students used an accommodation other than those listed on the NAEP questionnaire.

Table 2.13 — Percentage of students with disabilities receiving selected response accommodations used in achievement testing by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD NAEP

Types of response accommodation	Grade 4	Grade 8	Grade 12
Oral responses	14	8	8
Use of Braille/talking calculator	1	9	10
Pointing to answers	5	2	2
Use of computer to respond	2	2	5
Tape recording of answers	2	0	1
Use of typewriter	1	0	0
Accommodation not listed on questionnaire *	4	2	5

* Bilingual test booklet was not offered at grade 12.

NOTE: Questionnaire respondents were instructed to choose all responses that apply.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Chapter Summary

This chapter provided an overview on the demographic characteristics, nature and severity of disabilities, and instructional experiences of the 1996 NAEP population of students with disabilities. The information presented was obtained from a questionnaire filled out for each student with a disability by a knowledgeable staff member. The results presented are based on the 1996 national NAEP mathematics sample.

The 1996 NAEP assessment collected some information on the nature and severity of student disabilities. For each student with a disability in the NAEP sample, questionnaire recipients were presented a list of disabilities and asked to indicate all those that applied. At all three grades the vast majority of students with disabilities was classified into at least one of four of the listed categories (learning disability, speech/language impairment, mental or cognitive impairment, and emotional disturbance). Learning disability was the most frequently reported category with close to three of four students so identified at each of the three grades.

Recipients of NAEP's SD/LEP questionnaire were asked to indicate the degree of each student's disability with respect to four possible categories: mild, moderate, severe, or profound. About half of the students at each grade was described as having mild disabilities. The remaining half at each grade was almost all categorized with moderate to severe disabilities. Very few students who received educational services at schools participating in NAEP (1 percent each at grades four and eight, and 3 percent at grade 12) were judged to have profound disabilities. Severe or profound disabilities appeared to be more common among students with cognitive or mental impairments than among students with other disabilities.

SD/LEP questionnaire recipients were asked to indicate for each student with a disability the percentage of time spent mainstreamed in academic subjects. Regardless of grade level, about half of all students with disabilities were mainstreamed in academic subjects at least 80 percent of the time. At the other end of the continuum, the use of self-contained classrooms for academic instruction appeared to be more prevalent at grades 8 and 12 than at grade 4.

In order to collect information on the level of instruction provided to students with disabilities, SD/LEP questionnaire recipients were asked to indicate the grade level of instruction each student was receiving in three distinct areas: reading/language arts, mathematics, and science. In reading/language arts, half or fewer of the students with disabilities received instruction that was at grade level. In mathematics and science, the situation was slightly better at the two lower grades. More than half of the grade 4 and grade 8 students with disabilities received grade-level instruction in mathematics, and near 70 percent of these students received grade-level instruction in science. The situation at grade 12 in mathematics and science was not so encouraging. In all three subject areas, students with disabilities at higher grades appear to be further behind with respect to grade level of instruction.

In all three subject areas and at all three grades, the majority of students with disabilities was reported to be receiving the same content of instruction as their nondisabled peers at the same grade level. However, at all three grades the situation differed dramatically for students at or above grade level versus those below grade level. In all three subject areas, almost all students who received instruction that was at or above grade level received the same curriculum content as their nondisabled peers. In contrast, with one exception, fewer than half of those students with disabilities who received below grade-level instruction was taught the same curriculum content as their nondisabled peers.

Most students with disabilities have IEPs that delineate the academic, behavioral, vocational, and social skills domains in which they will receive special education services. SD/LEP questionnaire recipients were presented with a list of seven areas, plus an "other" option and asked to indicate in which of these areas each student received instruction as part of the special education program. For students with disabilities, at least one area of special education instruction was indicated for 94 percent of the fourth-grade students, 87 percent of eighth-grade students, and 80 percent of twelfth-grade students. At all three grades, the most common areas of special education instruction were language development, reading, and mathematics. With one or two exceptions, special instruction in the other areas was much less common among students with disabilities. As might be expected, vocational education was almost nonexistent at the fourth grade, but appears to be more prevalent at the upper grades. At grade twelve, 38 percent of the students with disabilities received some form of vocational education as part of their school's special education programs.

As part of the 1996 NAEP assessments, SD/LEP questionnaire recipients were asked to indicate for each student the grade level at which he or she was performing in three academic areas: reading/language arts, mathematics, and science. The questionnaire results are consistent with the existing research in suggesting low levels of academic achievement for the majority of these students. In all three grades, about 75 percent of students with disabilities were judged to be performing below grade level in reading/language arts. Reported performance levels in mathematics and science appear somewhat higher than those in reading/language arts at grade 4. However, at the higher grades, a greater incidence of below-grade-level performance was reported.

One of the approaches being used to increase participation of students with disabilities in assessments is allowing accommodations or adaptations. Across the three grades, respondents reported that 42 to 44 percent of students with disabilities were granted some form of accommodation or adaptation in testing. Between 33 and 41 percent of students with disabilities used presentation, timing, or setting accommodations. Response accommodations were used less frequently, ranging from 15 percent at grade eight to 19 percent at grade twelve. Reading directions aloud, reading problems aloud, and providing assistance with directions were the most commonly reported presentation accommodations. Extended testing time was the most commonly reported timing accommodation. The most typical setting accommodation was testing students in small groups. None of the listed response accommodations was used with great frequency among students with disabilities.

Chapter 3

The Inclusion of Students with Disabilities in NAEP: A Closer Look

As discussed in chapter 1 of this report, the 1996 NAEP was designed to allow the introduction and evaluation of modified procedures intended to increase the participation rates in NAEP of students with disabilities. Modifications were made in two specific areas. First, inclusion criteria for the 1996 assessments were revised with the intention of making them clearer, more inclusive, and more likely to be applied consistently across jurisdictions participating in the state NAEP program. Second, assessment accommodations were offered to students with disabilities who are regularly tested with them and whose Individualized Educational Plans (IEPs) specified such accommodations.

Initial evaluations of the effect of these modifications on inclusion rates for students with disabilities were presented in the *NAEP 1996 Mathematics Report Card for the Nation and the States* [1] and the *NAEP 1996 Science Report Card for the Nation and the States*[2]. The evaluations indicated the following:

- The introduction of revised inclusion criteria, without the provision of accommodations, had at most a limited effect on the percentage of students with disabilities who were assessed in NAEP. At grade 4, it appeared from national results that *fewer* students with disabilities were assessed using the revised criteria – a result contrary to expectations and to the intentions of the revision. However, state NAEP results for the same grade did not corroborate this apparent finding. Moreover, at grades 8 and 12, results from both state and national NAEP assessments indicated similar inclusion rates and no significant differences in the percentages of students with disabilities who were assessed. Thus, on balance, the weight of the evidence suggests no effects associated with the revision of the inclusion policy in the absence of accommodations.
- The provision of accommodations increased the percentages of assessed students with disabilities at grade 4 and, to a lesser extent, at grade 8. Grade-4 national NAEP results

[1] Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). *NAEP 1996 mathematics report card for the nation and the states*. Washington, DC: National Center for Education Statistics.

[2] O'Sullivan, C.Y., Reese, C.M., & Mazzeo, J. (1997). *NAEP 1996 science report card for the nations and states.* Washington, DC: National Center for Education Statistics.

for both the mathematics and science assessments and grade-8 results in mathematics indicated that more students with disabilities were assessed when accommodations were made available. Inclusion rates with and without accommodations did not differ significantly for grade 8 science or for either subject at grade 12. It is worth noting however that, in each of these instances, the observed percentages assessed when accommodations were available were in the expected direction (i.e., higher, though not significantly so) than when accommodations were not available.

- Mathematics results from all three grades and science results from grades 4 and 8 indicated that fewer students with disabilities were assessed under standard conditions when accommodations and adaptations were made available. These results suggest that some students with disabilities will be assessed with accommodations when these are available, but will be assessed under standard conditions when special administration procedures are not available. This may indicate that accommodations may not be needed for some of these accommodated students or that some students who can be tested without accommodations may be better able to show what they know and can do with accommodations. In either case, the effect of this "switching" phenomenon on trend measurement and the development of procedures to minimize its impacts are important areas for continuing research and development.

This chapter presents results that look more closely at NAEP inclusion rates among students with disabilities. In the first section of the chapter, results originally included in the *NAEP Report Cards* are presented again and discussed in light of additional information available from the questionnaire to which school officials responded for students with disabilities/limited English proficiency. In the second section of this chapter, this diversity is explored further by examining inclusion rates among particular segments of the population of students with disabilities. An additional purpose of this section is to provide a more complete understanding of the impact of the revised inclusion criteria and the provision of accommodations among students with disabilities.

Assessing the Impacts of Criteria Revisions and the Provision of Accommodations

When considering the inclusion of students with disabilities in large-scale assessments, researchers at the National Center on Educational Outcomes[3] distinguish between three groups of students: those who are capable of taking the assessment without accommodation; those who are capable of taking the assessment with accommodations; and, those who will need to take a different assessment. According to one estimate,[4] the first two groups represent about 85 percent of all students with disabilities. It is argued that students in these two groups are, for

---

[3] Elliot, J., Thurlow, M.L., & Ysseldyke, J.E. (1996). *Assessment guidelines that maximize the participation of students with disabilities in large-scale assessments: Characteristics and considerations.* Minneapolis, MN: University of Minnesota, National Center for Educational Outcomes.

[4] Ysseldyke, J.E., Thurlow, M.L., McGrew, K.S., & Shriner, J.G. (1994). *Recommendations for making decisions about the participation of students with disabilities in statewide assessment programs.* (Synthesis Report No. 15). Minneapolis, MN: University of Minnesota, National Center for Educational Outcomes.

the most part, exposed to a general education curriculum and hence, such students, with accommodations, should be included in large-scale assessments based on that curriculum.

Table 3.1 presents the percentages of students with disabilities who were assessed under standard conditions, assessed with accommodations, and excluded from the assessment for each of the three national samples. These percentages were originally included in a slightly different format in the *NAEP 1996 Mathematics Report Card.*[5]

For the samples in which accommodations were not offered (i.e., S1 and S2), the percentages of assessed students with disabilities ranged from 47 to 58 percent. These inclusion

Table 3.1 — Percentage of students with disabilities in the national population included in the NAEP assessment, by grade and sample type: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD

	Student participation in NAEP				
		Assessed			
Sample type	N	% assessed without accommodations	% assessed with accommodations	Total % assessed	% excluded
Grade 4					
S1[1]	359	58	—	58	42
S2[2]	411	47	—	47	53
S3[3]	424	35† ††	37	72† ††	28† ††
Grade 8					
S1[1]	310	55	—	55	45
S2[2]	524	58	—	58	42
S3[3]	557	46††	26	71† ††	29† ††
Grade 12					
S1[1]	211	48	—	48	52
S2[2]	411	51	—	51	49
S3[3]	386	35††	19	54	46

1 1990-1996 operational inclusion criteria

2 1996 revised inclusion criteria

3 Revised inclusion criteria plus accommodations in testing

† Indicates a significant difference between S1 and S3 results.

†† Indicates a significant difference between S2 and S3 results.

NOTE: — Not applicable because accommodations were not offered

N's in Table reflect only students for whom matching background questionnaire data were available.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

---

5 Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). *NAEP 1996 mathematics report card for the nation and the states*. Washington, DC: National Center for Education Statistics.

percentages, which are consistent with those from previous NAEP assessments where accommodations were not offered, represent a level of participation for students with disabilities that is substantially less than that which Ysseldyke and his colleagues suggest should be possible. When accommodations were available (i.e., in the S3 samples), the percentages of assessed students with disabilities ranged from 54 to 72 percent – values that indicate improvement (statistically significant in grades 4 and 8), but still represent levels below those some think should be possible.

Responses to several questions on the NAEP SD/LEP questionnaire provide an additional context within which to discuss NAEP inclusion rates, with and without the provision of accommodations. Questionnaire recipients were asked whether or not their students with disabilities could meaningfully participate in NAEP without accommodations or adaptations. Table 3.2 presents results based on this question for the combined mathematics sample. It is of particular interest to compare the percentages in table 3.2 with those in table 3.1.

Table 3.2 — Percentage of students with disabilities who could meaningfully participate in NAEP mathematics without accommodations or adaptations by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Could this student meaningfully participate in NAEP without accommodations?	Yes	No
Grade 4	38	62
Grade 8	57	43
Grade 12	43	57

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

At grade 8, questionnaire recipients indicated that 57 percent of students with disabilities could meaningfully participate in NAEP without accommodations. These percentages match the table 3.1 percentages of assessed students for S1 and S2 fairly well. The percentages assessed without accommodation in S3 were about 10 points lower, suggesting that when accommodations are available, some students may receive them despite judgments that meaningful measurement may be possible without them. At grade 12, questionnaire recipients indicated that 43 percent of students with disabilities could participate meaningfully without accommodations. This percentage appears somewhat lower than that obtained in S1 and S2 (where accommodations were prohibited), but somewhat higher than obtained in S3 (where accommodations were permitted). Again, the results are consistent with the concern that some students who can meaningfully participate in NAEP without accommodations will be tested with accommodations when they are available. As noted earlier, it may be that students need accommodations to most effectively show their skills, but they can still perform without them (albeit, at potentially depressed levels). On the other hand, the grade 8 and grade 12 results could also indicate a tendency to provide accommodations in situations where they are not truly necessary.

The questionnaire results from grade 4, however, do not match up with inclusion percentages in quite the same way. At grade 4, questionnaire respondents indicated that only 38 percent of students with disabilities could meaningfully participate in NAEP without accommodations. This estimate is quite close to the percentage of students assessed without accommodations in the S3 sample (35 percent). The percentages assessed under standard conditions in the S1 and S2 samples, for which no special testing conditions were offered, were 10 to 20 percentage points higher than the questionnaire results might suggest.

The lack of a consistent relationship across the grades between questionnaire results and actual inclusion rates is disquieting and makes interpretation of any patterns somewhat difficult. Inconsistencies may suggest a lack of validity for questionnaire responses, the presence of inappropriate inclusion decisions, or possible response biases due to missing questionnaire data. Such inconsistencies may also reflect different inclusion tendencies and different practices in
the application of accommodations across the grades. Despite the differing patterns, there is one consistent aspect of the results. Where accommodations and adaptations were not provided, the percentages of students with disabilities tested in NAEP equaled or exceeded the percentages of students judged as capable of meaningfully participating without accommodations. ***Consequently, increases in the numbers of students with disabilities participating in NAEP are not likely to result solely from revisions to inclusion criteria that do not involve the provision of accommodations.*** In this respect, the results in tables 3.1 and 3.2 reinforce those reported in the *NAEP 1996 Report Cards* and in the *NAEP 1995 Field Test Report.*[6]

A second set of SD/LEP questionnaire items provides a somewhat different perspective from which to evaluate current levels of inclusion and the degree of improvement that might be attainable. Questionnaire recipients were asked to indicate for each student the conditions under which the student would participate in NAEP if accommodations/adaptations were available as follows: 1) without accommodations or adaptations; 2) with the accommodations or adaptations specified for achievement testing with this student; and 3) the student cannot participate in assessments such as NAEP as determined by the IEP team or an equivalent group. Results based on the combined sample from the mathematics assessment are presented in table 3.3.

---

[6] Anderson, N.E., Jenkins, F.F., & Miller, K.E. (1996). *NAEP inclusion criteria and testing accommodations: Findings from the NAEP 1995 field test in mathematics.* Princeton, NJ: Educational Testing Service.

Table 3.3 — Percentage distribution of participation status for students with disabilities by grade: NAEP 1996 mathematics sample

If accommodations or adaptations were available, how would this student participate in NAEP?	Without accommodations	With accommodations	Would not participate
Grade 4	27	43	30
Grade 8	38	41	21
Grade 12	29	38	34

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Questionnaire respondents indicated that 30 percent of grade 4 students, 21 percent of the grade 8 students and 34 percent of grade 12 students with disabilities could not participate in NAEP. The results in table 3.1 for the S3 sample show that for grades 4, 8, and 12, the percentages of students with disabilities excluded from NAEP were 28, 29 and 46, respectively. Thus, while the grade 4 results suggest that the permitted accommodations and adaptations may have been sufficient to include everyone who was not explicitly excluded from participation on the basis of their IEP, ***the grade 8 and grade 12 results suggest that further modest improvements in inclusion are still possible.***

Another interesting aspect of the results in table 3.3 is evident in the reported percentages of students who would participate in NAEP with accommodations. These percentages (43 percent at grade 4, 41 percent at grade 8, and 38 percent at grade 12) appear higher than the percentages actually assessed with accommodations in the S3 samples (37 percent at grade 4, 26 percent at grade 8, and 19 percent at grade 12, respectively). Moreover, at all three grades, the percentages of students assessed without accommodations (table 3.1) appear higher than the percentages suggested by the questionnaire responses (table 3.3). Both these sets of results are consistent with the notion that an expansion of accommodations or adaptations permitted by NAEP, or a change in NAEP guidelines regarding eligibility for special testing conditions, could result in further increases in inclusion rates.

For reasons of cost and feasibility, NAEP did not attempt to provide all possible accommodations and adaptations in all grades and subjects. Nor did participating schools offer all students with disabilities the opportunity to test with accommodations or adaptations. Eligibility for accommodations or adaptations in NAEP had to be specified in a student's IEP or be routinely provided by the school in other testing situations. For the most part, NAEP permitted any and all accommodations usually provided by the school that would not interfere with the intent of the assessment. The adaptations offered included braille booklets (at grades 8 and 12), large print booklets (at all grades), and a Spanish bilingual booklet (at grades 4 and 8, in mathematics). Table 3.4 presents results from the mathematics assessment on the relative frequency of the use of various accommodations and adaptations among S3 students receiving nonstandard administrations.

Table 3.4 — Percentage of students with disabilities assessed in NAEP with each offered accommodation type by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Accommodation/adaptation type	Grade 4	Grade 8	Grade 12
Small group session	17	11	7
Extended time (regular session)	8	7	5
One-on-one testing	8	4	4
Directions read aloud (regular session)	4	2	3
Bilingual test booklet	0	0	0
Bilingual dictionary	0	0	0
Braille, large type booklet	0	0	0
Other accommodations	0	1	1

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Essentially all nonstandard administrations involving students with disabilities used one or more accommodations. At all three grades the most common accommodation involved testing in small groups. Such sessions, which frequently involved extended time limits and the reading of directions and/or questions aloud, were used with 17 percent, 11 percent and 7 percent of accommodated students at grades 4, 8, and 12, respectively. Some students with disabilities (in particular those with physical disabilities) may need to take tests in a one-on-one setting. Such testing is typically carried out under extended time limits and with the aid of a facilitator to read questions and/or record responses. One-on-one testing was carried out for between 4 and 8 percent of the accommodated students across the three grades. The use of extended time limits within regular NAEP testing sessions was also a common accommodation, involving between 5 and 8 percent of accommodated students. Some students within regular testing sessions required particular words, phrases, or sentences to be read to them.
Such accommodations were made for 2 to 4 percent of accommodated students across the three grades.

It is interesting to note that among students with disabilities, almost no use was made of the offered adaptations, in particular the braille and large-type versions of the tests. Apparently, the relatively small percentages of students with visual impairments that were encountered in the NAEP samples were either excluded from the assessment or tested with some other accommodation. Thus, it is difficult to conjecture about what else the NAEP program could offer in the way of adaptations that would materially extend inclusion rates among students with disabilities.

Within the bounds of practical realities, fiscal constraints, and evolving policies, the NAEP program has always sought to include in its assessments as many students with disabilities as possible. However, NAEP inclusion procedures and policies also need to reflect the voluntary nature of the program and be sensitive to the concerns and experiences of the schools and students that participate. In this regard, assessments have been carried out in a manner intended to encourage the inclusion of all students except in the following prespecified situations: (1) a student's IEP explicitly indicates exemption from testing in assessments such as NAEP; (2) significant cognitive disabilities exist that make the student incapable of participating; or (3) the necessary accommodations or adaptations are unavailable.

The SD/LEP questionnaire contained two questions that allow some evaluation of the degree to which inclusion decisions are made on the basis of the intended program criteria. The first question asked whether the IEP team or equivalent group had determined that the student could not participate in NAEP. A second question asked whether the student's cognitive functioning was so severely impaired that participation was not possible. Table 3.5 presents the percentages of excluded students broken down by respondents' answers to these two questions.

Table 3.5 — Percentage distribution of reasons for exclusion of students with disabilities from NAEP assessment, by sample type and grade: NAEP 1996 mathematics sample

Sample Type	N	Reason for Exclusion: Stated in IEP and judged to be impaired	Stated in IEP (only)	Judged to be impaired (only)	Neither reason
Grade 4					
S1[1]	143	58	18	3	22
S2[2]	190	59	18	1	22
S3[3]	80	64	21	2	13
Grade 8					
S1[1]	112	50	22	6	22
S2[2]	161	45	7 ‡	6	43 ‡
S3[3]	152	33	30††	10	27
Grade 12					
S1[1]	89	52	12	6	30
S2[2]	199	51	16	3	31
S3[3]	146	60	16	2	22

1 1990-1996 operational inclusion criteria

2 1996 revised inclusion criteria

3 Revised inclusion criteria plus accommodations in testing

‡ Indicates a significant difference between S1 and S2 results.

†† Indicates a significant difference between S2 and S3 results.

NOTE: Because of rounding, percentages may not add to 100. Results in this table are based on students for whom matching background questionnaire data were available. As a result, sample sizes do not match those given in Table 1.2.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

One possible reason for these additional exclusions is the absence in NAEP of a needed accommodation adaption regularly used by such students. The fact that, in all but one instance, the percentages of students excluded for neither reason appears lower in the S3 sample (where accommodations were permitted) than in the other samples (where accommodations were not permitted) — is consistent with such a conjucture. However, the fact that, even in S3, 27 percent of exclusions were based neither on IEP's nor on judgment of severe impairement suggest that other factors may be at work here.

There are a number of aspects of these exclusion results that are worth noting. First, it is apparent that, for most students, decisions about their exclusion from the assessment are made on the basis of what is stated in their IEP. Across sample types and grades, between 52 and 85 percent of exclusions involved students whose IEP stated that they were not to be tested. In all but one instance, the percentage of students excluded for this reason exceeds 63 percent. Furthermore, relatively few exclusion decisions, 6 percent or less in all but one instance, appear to be made on the basis of severe cognitive impairment, in the absence of direction from the IEP. However, the results in table 3.5 show that for substantial percentages of excluded students (between 13 and 43 percent, across samples and grades, and under 31 percent in all but one instance), neither determination by the IEP team nor the presence of cognitive impairments was given as the reason for exclusion. This result could indicate that respondents were not filling out the questionnaire properly. However, other factor could be at work. Some of these exclusion particularily in the S1 and S2 samples, could be due to the unavailability of necessary accommodations and adaptations. A 1997 NCFO[7] report also suggests a number of other factors that may work against the full participation of students with disabilities in large scale assessment programs. These factors include a desire on the part of parents, teachers, and others to protect such students from "stressful" situations. However, other factors could be also at work. Some of these exclusions, particularly in the S1 and S2 samples, could be due to the unavailability of necessary accommodations and adaptations. Additional research into what these other reasons for exclusion might be could benefit the program in crafting future changes to policies and procedures, to encourage fuller participation.

A Look at Inclusion Rates within Selected Subgroups

Level of Disability and Inclusion Rates. As noted in chapter 2, the vast majority of students with disabilities who are encountered in NAEP-eligible schools have a moderate or mild degree of disability. In assessing the impacts of the revised procedures on inclusion in NAEP, it is of interest to examine inclusion rates from the three samples (S1, S2, and S3) by degree of disability. Results of this analysis appear in table 3.6.

As might be anticipated, the data show an apparent association between degree of disability and rates of inclusion in NAEP at all three grades. Regardless of sample type, inclusion rates appear highest among students with mild disabilities and lowest among students with severe or profound disabilities. ***At all three grades, the revisions to the inclusion***

---

[7] National Center on Educational Outcomes. (1997). *1997 state special education outcomes: A report on state activities during educational reform.* Minneaapolis, MN: Author.

Table 3.6 — Percentage of students included in NAEP by degree of disability, by grade and sample type: NAEP 1996 mathematics sample

What is the degree of this student's disability?	Mild			Moderate			Severe/Profound		
	N	% assessed	% excluded	N	% assessed	% excluded	N	% assessed	% excluded
Grade 4									
S1[1]	151	74	26	112	45	55	48	24	76
S2[2]	153	59	41	136	37	63	52	27	73
S3[3]	160	84 ††	16 ††	117	69 † ††	31 † ††	22	37	63
Grade 8									
S1[1]	117	78	22	78	55	45	38	6	94
S2[2]	199	79	21	122	59	41	58	16	84
S3[3]	207	88	12	173	63	37	58	45 † ††	55 † ††
Grade 12									
S1[1]	78	64	36	63	43	57	25	32	68
S2[2]	159	66	34	109	48	52	60	9	91
S3[3]	130	64	36	99	50	50	36	25	75

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

† Indicates a significant difference between S1 and S3 results.

†† Indicates a significant difference between S2 and S3 results.

NOTE: Because of rounding, percentages may not add to 100.

N's in table reflect only students for whom matching background questionnaire data were available.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

criteria, absent the provision of accommodations, resulted in no significant differences in inclusion percentages regardless of the degree of student disabilities. However, the provision of accommodations did result in significantly increased inclusion percentages in grades 4 and 8 in a number of instances. The pattern of these increases suggests that improvements at these two grades were not restricted to a particular degree of disability.

Grade 4 inclusion percentages were higher in S3 (where accommodations were provided) than in S2 (where the same revised inclusion criteria were used, but accommodations were not provided) for students with mild and moderate disabilities. Grade 4 inclusion percentages were also higher in S3 than in S1 (where the original inclusion criteria were used and no accommodations offered) among students with moderate disabilities. Among grade 4 students with severe or profound disabilities, the results follow a similar pattern, with S3 inclusion percentages appearing higher than those in S1 or S2. However, these apparent differences are not statistically significant. At grade 8, inclusion percentages in S3 were higher than those in S1 and S2 among students with severe or profound disabilities. Among grade 8 students with mild and moderate disabilities, results again follow the pattern of S3 inclusion percentages, appearing to be higher than those in S1 or S2. However, these apparent differences are not statistically significant. At grade 12, no significant differences in inclusion rates were noted.

Time Mainstreamed and Inclusion Rates. As noted in chapter 1, one of the purposes of revising NAEP's inclusion criteria was to encourage inclusion decisions for students regardless of the percentage of time they are mainstreamed. The original inclusion criteria stated that students with disabilities could be excluded if they were mainstreamed in academic subjects less than 50 percent of the time. In the revised criteria, no mention was made of percentage of time mainstreamed. Students were to be included unless their IEP stated they were not to be tested or a staff member most knowledgeable about the student judged that he or she could not meaningfully participate in the assessment. Given the pivotal role of percentage of time mainstreamed in the revisions to the inclusion criteria, it is of some interest to examine inclusion percentages for each of the three samples by percentage of time mainstreamed.

Table 3.7 shows the percentages of students with disabilities assessed in NAEP and the percentage excluded by sample type for two groups of students: 1) those mainstreamed in academic subjects less than 50 percent of the time, and 2) those mainstreamed in academic subjects 50 percent or more of the time. As might be expected, inclusion rates appear noticeably higher among students mainstreamed in academic subjects 50 percent or more of the time, regardless of grade or sample type.

Table 3.7 — Percentage of students with disabilities included in NAEP by percentage of time mainstreamed, by sample type and grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

What percentage of time is this student mainstreamed in academic subjects?	Less than 50%			50% or more		
	N	% assessed	% excluded	N	% assessed	% excluded
Grade 4						
S1[1]	68	15	85	243	66	34
S2[2]	90	16	84	253	55	45
S3[3]	67	44 † ††	56 † ††	235	80 † ††	20 † ††
Grade 8						
S1[1]	76	23	77	164	77	23
S2[2]	154	43	57	233	76	24
S3[3]	173	50 †	50	275	85	15
Grade12						
S1[1]	46	20	80	126	62	38
S2[2]	137	18	82	193	67	33
S3[3]	111	32	68	156	68	32

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

† Indicates a significant difference between S1 and S3 results.

†† Indicates a significant difference between S2 and S3 results.

NOTE: Because of rounding, percentages may not add to 100.

N's in table reflect only students for whom matching background questionnaire data were available.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Comparisons of inclusion percentages for the S1 and S2 samples by percentage of time mainstreamed corroborate the overall results in suggesting that the revisions to the inclusion criteria, absent the provision of accommodations and adaptations, had little measurable effect. At all three grades, there were no statistically significant differences between S1 and S2 inclusion percentages. Moreover, no consistent pattern of results was apparent across grades.

Comparisons of the S3 inclusion rates with those of S1 and S2 suggest that at grade 4 there were increases in inclusion resulting from the provision of accommodations and adaptations for grade 4 students both above and below the 50 percent mainstreamed criterion. Grade 4 inclusion percentages were higher in S3 than in S1 or S2 for both groups of students. At grade 8, inclusion percentages also appear higher in S3 than in S1 or S2 for both groups of students. However, only one of these apparent differences (S3 versus S1 for students mainstreamed less than 50 percent of the time) is statistically significant. At grade 12, no statistically significant differences were obtained.

Above/below Grade Level and Inclusion Rates. As noted in chapter 2, some students with disabilities, particularly at grades 4 and 8, are receiving mathematics instruction that is at grade level and, for the vast majority of these students, is comparable to that received by their nondisabled peers. Other students with disabilities are receiving instruction that is below their grade level and, in some cases, reflects a different curriculum content than that presented to their nondisabled peers. Table 3.8 shows the percentage of students included in NAEP for three groups of students with disabilities: students receiving grade-level instruction; students receiving below-grade-level instruction with the same curriculum content as their nondisabled peers; and students receiving below-grade-level instruction and different curriculum content than their nondisabled peers. As might be expected, the general pattern of results suggests that at all three grades, students receiving instruction at grade level appeared to be included at higher rates than those receiving below-grade-level instruction.

As with a number of the analyses presented here, comparisons between the sample types suggest that with one exception , the provision of accommodations and adaptations at grades 4 and 8 appeared to increase inclusion for all three groups of students. At grade 4, inclusion percentages were significantly higher in S3 than in S2 for all three groups of students. S3 inclusion percentages were also higher than those obtained in S1 for grade 4 students receiving grade-level instruction. S3 inclusion percentages were higher than those obtained in S1 for grade 8 students receiving grade-level instruction and those receiving below-grade-level instruction and different curriculum content than their nondisabled peers. Grade 8 inclusion rates were also higher in S3 than in S2 for students receiving below-grade-level instruction and different curriculum content than their nondisabled peers.

The results at grade 12 are complicated and do not follow a consistent pattern of increased inclusion under S3 conditions. Among students receiving grade-level instruction, inclusion percentages in S3 were significantly lower than those in S1 and S2, a result contrary to expectation. In contrast, S3 inclusion percentages were higher than those in S1 and S3 among students receiving both below-grade-level instruction and a different curriculum than their nondisabled peers. It should be noted that the percentage of grade 12 students with disabilities receiving grade-level instruction is somewhat small and the result for this group may be quite unstable. No other explanations for this odd pattern are readily apparent.

Table 3.8 — Percentage of students with disabilities included in NAEP by grade level of instruction and curriculum content in mathematics, by sample type and grade: NAEP 1996 mathematics sample

	At/above grade level			Below grade level					
	All students			Students with same curriculum as non-disabled students			Students with different curriculum from nondisabled students		
Sample Type	N	% assessed	% excluded	N	% assessed	% excluded	N	% assessed	% excluded
Grade 4									
S1[1]	175	69	31	53	36	64	77	30	70
S2[2]	169	66	34	81	26	74	72	11	89
S3[3]	168	90 † ††	10 † ††	52	60 ††	40 ††	74	44 ††	56 ††
Grade 8									
S1[1]	104	77	23	50	68	32	83	30	70
S2[2]	180	82	18	61	73	27	130	30	70
S3[3]	250	89 †	11 †	72	56	44	118	51 † ††	49 † ††
Grade 12									
S1[1]	43	87	13	50	68	32	51	10	90
S2[2]	78	84	16	46	43	57	127	18	82
S3[3]	58	65 † ††	35 † ††	52	47	53	106	42 † ††	58 † ††

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

† Indicates a significant difference between S1 and S3 results.

†† Indicates a significant difference between S2 and S3 results.

NOTE: Because of rounding, percentages may not add to 100.

N's in table reflect only students for whom matching background questionnaire data were available.

At all grades, the at/above grade level category includes students with "same" curriculum (95%) and "different curriculum" (5%) as nondisabled students.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Accommodations and Inclusion Rates. As noted earlier, not all students with disabilities in the S3 sample were offered the opportunity to test with accommodations or adaptations. In order to be eligible for accommodations or adaptations in NAEP, their use in testing was supposed to be specified in a student's IEP or be routinely provided by the school in other testing situations. As part of the SD/LEP questionnaire, respondents indicated for each student whether accommodations or adaptations are used for achievement testing. Three categories of response were permitted: yes; no; or IEP states that the student cannot be tested. Table 3.9 presents results on NAEP inclusion percentages for three categories of students: (1) students who receive accommodations or adaptations for achievement testing; (2) students who do not receive accommodations or adaptations for achievement testing; and (3) students whose IEP states they cannot be tested. For each group of students, table 3.9 presents the percentage routinely assessed without accommodations or adaptations, the percentage routinely assessed with accommodations or adaptations, and the percentage excluded from the assessment. Results are presented separately for the S1, S2, and S3 samples at all three grades.

A number of interesting and important patterns are evident in these results. First, one might conjecture that the impact of providing accommodations and adaptations should be most evident among students reported to be receiving accommodations or adaptations (the "yes" group). That is, NAEP inclusion rates should be highest for this group (or equivalently, NAEP exclusion rates should be lowest for this group) in the S3 sample, where accommodations and adaptations were permitted. The results provide some, but not total, support for this conjecture. The percentage of students excluded from NAEP appeared lower in S3 than in S1 or S2 at all three grades. However, of these apparent differences, only the S2/S3 difference at grade 4 and the S1/S3 difference at grade 8 were statistically significant. Second, one might conjecture that, ideally, the provision of accommodations and adaptations should have limited impact on students reported not to be receiving accommodations (the "no" group). Again, the results in table 3.9 provide some support for this conjecture. It is indeed the case that among students reported not to be receiving accommodations or adaptations, there were no significant differences between the sample types in the percentages of excluded students.

The results in table 3.9 also present inconsistencies that point to important methodological issues. It is evident that in the S1 and S2 samples (where accommodations and adaptations were not offered) substantial percentages of students reported to be receiving accommodations or adaptations in achievement testing were included in NAEP assessments and tested under standard conditions. In the S3 samples (where accommodations and adaptations were offered in NAEP) many, but not all, students were tested with accommodations or adaptations. For example, despite the availability of accommodations and adaptations, 20 percent of grade 4 students in the S3 sample, who ostensibly would receive accommodations or adaptations in achievement testing, were tested in NAEP under standard conditions. The corresponding percentages at grades 8 and 12 were 31 and 38 percent, respectively. It is unclear why, when special testing conditions are offered, students reported to be receiving accommodations or adaptations in achievement testing would be assessed in NAEP under standard conditions. One possibility is that the accommodations or adaptations routinely used were not readily available or easily implemented for the NAEP test. A second possibility is that incorrect decisions were made in the field regarding the conditions under which individual students should be assessed. A further possibility is that the individual filling out the SD/LEP questionnaire made an incorrect choice. Follow-up research in future assessments will be required to examine these possibilities further.

As noted earlier in this report, the intent of the NAEP program was to offer accommodations or adaptations only to those students who regularly test under special conditions. The S3 results in table 3.9 for students who do not receive accommodations and adaptations suggest that this intent may not have been fully realized. At each grade, some percentage of S3 students not routinely using accommodations or adaptations in achievement testing were tested in NAEP with an accommodation or adaptation. The percentages ranged from 8 to 12 percent. Moreover, students with IEPs that state they cannot be tested are expected to be excluded from the NAEP samples. However, the results in table 3.9 suggest that this may not always have been the case. At all grades and in all samples, some percentage of students reported to have IEPs that state they cannot be tested were included in the NAEP sample. In some cases, the testing apparently occurred under standard conditions and in other instances the testing was carried out using an accommodation or adaptation. Again, the reason for such apparent inconsistencies is not clear and could be due, among other things, to incorrect decisions made in the field or invalid questionnaire data.

Table 3.9 — Percentage of students with disabilities assessed in 1996 mathematics by accommodation/adaptation status by sample type and grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Are accommodations/ adaptations used for achievement testing for this student?	Yes				No				IEP says student cannot be tested			
	Participation in NAEP				Participation in NAEP				Participation in NAEP			
	N	% assessed standard	% assessed with accom.	% excluded	N	% assessed standard	% assessed with accom.	% excluded	N	% assessed standard	% assessed with accom.	% excluded
Grade 4												
S1[1]	97	61	—	39	131	76	—	24	80	11	—	89
S2[2]	112	43	—	57	136	74	—	26	89	6	—	94
S3[3]	143	20	63	17††	101	82	8	11	56	6	13	81
Grade 8												
S1[1]	75	56	—	44	122	78	—	22	40	3	—	97
S2[2]	158	66	—	34	153	86	—	14	71	22	—	78
S3[3]	183	31	48	21 †	181	76	12	12	81	10	5	85
Grade 12												
S1[1]	51	57	—	43	73	69	—	31	45	4	—	96
S2[2]	135	57	—	43	88	72	—	28	99	15	—	85
S3[3]	94	38	35	27	86	65	8	27	85	2	1	97††

NOTE: — Not applicable because accommodations were not offered

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

† Indicates a significant difference between S1 and S3 results.

†† Indicates a significant difference between S2 and S3 results.

NOTE: Because of rounding, percentages may not add to 100.

N's in table reflect only students for whom matching background questionnaire data were available.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Chapter Summary

This chapter presents a closer look at NAEP inclusion rates among students with disabilities. In the first section of the chapter, results originally included in the *NAEP Report Cards* are presented and discussed in light of additional information available from the SD/LEP questionnaire. In the second section of this chapter, inclusion rates are examined among particular segments of the SD population. An additional purpose of this section is to provide a more complete understanding of the impacts of the revised inclusion criteria and the provision of accommodations on these particular segments, in particular those segments that one would have expected to be particularly affected by the changes made to inclusion criteria or to the provision of accommodations.

The results of the NAEP SD/LEP questionnaire provide an additional context within which to discuss NAEP inclusion rates, with and without the provision of accommodations. Questionnaire recipients were asked whether or not students could meaningfully participate in NAEP without accommodations or adaptations. ***Where accommodations and adaptations were not provided, the percentages of students with disabilities tested in NAEP equaled or exceeded the percentages of students judged as capable of meaningfully participating without accommodations. Consequently, increases in the numbers of students with disabilities participating in NAEP are not likely to result solely from revisions to inclusion criteria that do not also involve the provision of accommodations.***

Questionnaire recipients were asked to indicate for each student the conditions under which that student would participate in NAEP if accommodations and adaptations were available: (1) without accommodations or adaptations; (2) with the accommodations or adaptations specified for achievement testing with this student; or (3) the IEP team or equivalent group has determined that the student cannot participate in assessments such as NAEP. ***A comparison of the results from this question with the actual participation rates from the 1996 mathematics assessment suggests that an expansion of accommodations or adaptations permitted by NAEP, or a change in NAEP guidelines as to who is eligible for special testing conditions, could result in further small increases in inclusion rates.***

Current and previous NAEP assessments have been carried out in a manner intended to encourage the inclusion of all students except in certain prespecified situations: (1) a student's IEP explicitly indicates exemption from assessments such as NAEP; (2) significant cognitive disabilities exist that make the student incapable of participating; or (3) the necessary accommodations or adaptations are unavailable. An analysis of the reasons given for exclusion suggest that most exclusion decisions were made on the basis of what is stated in the IEP and relatively few exclusion decisions were made on the basis of a judgment of severe cognitive impairment, absent corroborating direction from the IEP. However, the results also suggest that for substantial percentages of excluded students, neither determination by the IEP team nor the presence of cognitive impairments was given as reason for exclusion. Additional follow-up research into what, if any, other reasons for exclusion are used could benefit the program in crafting future changes to policies and procedures.

The vast majority of students with disabilities who are encountered in NAEP-eligible schools have a moderate or mild degree of disability. However some students were encountered in the NAEP samples who were categorized by questionnaire respondents as having severe or profound disabilities. An examination of inclusion rates by degree of student disability confirmed that the revisions to the inclusion criteria, absent the provision of accommodations, resulted in no significant differences in inclusion percentages regardless of the degree of student disabilities. The provision of accommodations did result in significantly increased inclusion percentages in grades 4 and 8 in a number of instances, and the pattern of these increases suggests that improvements at these two grades were not restricted to a particular degree of disability.

As noted in chapter 1, one of the purposes of revising NAEP inclusion criteria was to encourage inclusion decisions for students regardless of the percentage of time they are mainstreamed. Comparisons of inclusion percentages for the S1 and S2 samples by percentage of time mainstreamed corroborate the overall results in suggesting that the revisions to the inclusion criteria, absent the provision of accommodations and adaptations, had no consistent measurable effect. Comparisons of the S3 inclusion rates with those of S1 and S2 suggest that, at grade 4, increases in inclusion resulting from the provision of accommodations and adaptations were evident for students both above and below the 50 percent mainstreamed criterion. At grade 8, results follow a similar pattern, but not all comparisons were statistically significant.

Some students with disabilities, particularly those at grades 4 and 8, received mathematics instruction at grade level and, for the vast majority of these students, the instruction was comparable to that received by their nondisabled peers. Other students with disabilities received instruction below their grade level which, in some cases, reflects a different curriculum content than that presented to their nondisabled peers. As might be expected, the general pattern of results suggests that at all three grades, students who received instruction at grade level appear to be included at higher rates than those receiving below-grade-level instruction. Comparisons between the sample types, with one exception, reveal that the provision of accommodations and adaptations at grades 4 and 8 appears to increase inclusion for the three groups of students. The results at grade 12 are more complicated and do not follow a consistent pattern of increased inclusion under S3 conditions.

As part of the SD/LEP questionnaire, respondents were asked to indicate for each student whether accommodations or adaptations are used for achievement testing. Three categories of response were permitted: yes; no; or IEP states that the student cannot be tested. A number of interesting and important patterns are evident in these results. One might conjecture that: (1) the impact of providing accommodations and adaptations should be most evident among students reported to have received accommodations or adaptations, and; (2) the provision of accommodations and adaptations should have limited impact on students reported not to have received accommodations. The results provide some, but not total, support for these conjectures. However, the results for this question also present inconsistencies that point to important methodological issues. In particular, they suggest that incorrect decisions regarding inclusion may have been made in the field or, in some cases, that invalid data may have been provided on the SD/LEP questionnaire. Follow-up research in future assessments will be required to resolve such inconsistencies.

Chapter 4

Limited English Proficient Students: A Description of the NAEP Population

According to the National Center for Education Statistics (NCES) 1993-94 Schools and Staffing Survey,[1] over 2.1 million public school students were identified as limited English proficient (LEP) students. Based on data from the 1996 NAEP assessments, LEP students made up 5 percent of the nation's fourth graders, 4 percent of the nation's eighth graders, and 2 percent of the nation's twelfth graders. The NAEP assessments also indicate that among fourth-grade LEP students, about 1 in 5 were also classified as students with disabilities. As shown in appendix A (table A5.b), LEP students are largely from Hispanic or Asian ethnic backgrounds. Many are from disadvantaged socioeconomic backgrounds, as evidenced by the substantial percentages of LEP students that were receiving Title I services and were eligible for federal free or reduced-price lunches (see tables A6.b and A7.b and August & Hakuta[2]).

This chapter presents a description of the nation's fourth-, eighth-, and twelfth-grade LEP students based on questionnaire results obtained as part of NAEP's 1996 assessment. Based on the 1996 state NAEP assessments, LEP students made up over 10 percent of all public-school fourth graders in five states (Arizona, California, Florida, New Mexico, and Texas), and over 10 percent of public-school eighth graders in four states (California, Florida, New Mexico, and Texas). LEP students are largely concentrated in the West, in urban areas, and in large schools with 750 or more students.[3] Survey data from the Office of Biligual Education and Minority Language Affairs (OBEMLA) also shows substantial increases in LEP students in other states that historically have had few, if any, such students.[4] Thus, the inclusion and assessment of LEP students is a becoming a concern in a growing number of states.

---

[1] Han, M., Baker, D., & Roderiguez, C. (1997). *A profile of policies and practices for limited English proficient students: Screening methods, program support, and teacher training.* (NCES Publication No. 97-472). Washington DC: National Center for Education Statistics.

[2] August, D. & Hakuta, K. (Eds.). (1997) *Improving schooling for language-minority children: A research agenda.* Washington, DC: National Research Council. Institute of Medicine. Commission on Behavioral and Social Sciences and Education.

[3] Han, M., Baker, D., & Roderiguez, C. (1997). *A profile of policies and practices for limited English proficient students: Screening methods, program support, and teacher training.* (NCES Publication No. 97-472). Washington DC: National Center for Education Statistics.

[4] Summary Report of the Survey of the States' limited English proficient students and available education programs and serves (1994-1995), (1996). Washington, DC: National Council for Biligual Education.

Return rates for the SD/LEP questionnaire (see page 11) based on the 1996 national mathematics samples ranged from 76 percent (at grade 4) to 68 percent (at grade 12), and comparable return rates were observed for the national science samples. Analyses, presented in appendix A, indicate only minor differences between the subset of LEP students with returned questionnaires and all sampled LEP students with respect to a number of school- and student-level variables. Based on these analyses, it is reasonable to assume that the results presented here and in chapter 5 are reasonably representative of the LEP population of fourth, eighth, and twelfth graders in the NAEP sample.

A knowledgeable school staff member (e.g., a bilingual education or ESL teacher) was asked to fill out the SD/LEP questionnaires for each LEP student, regardless of whether the student was judged capable of participating in the NAEP assessment. Table 4.1 presents information on the position of the person(s) who completed the questionnaires.

Table 4.1 — Percentage distribution of job titles of limited English proficient student questionnaire respondents by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD NAEP

Who filled out the questionnaire?	Grade 4	Grade 8	Grade 12
Bilingual education teacher	17	27	18
Classroom teacher	39	12	9
Special education teacher	2	4	3
Principal	0	5	2
Other	2	9	20
Missing	40	43	50

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

For 40 to 50 percent of students, the position of the person filling out the questionnaire was not indicated. Among the questionnaires that did provide responses to this question, answers for the fourth- and eighth-grade students were typically provided by the classroom teacher or the bilingual education/ESL teacher. The bilingual education/ESL teacher or someone other than those listed typically provided answers for twelfth graders. Anecdotal reports from the field suggest that the "other" individual filling out the questionnaire was typically a classroom aide working closely with the student or a school counselor (e.g., the school psychologist). The results presented in this chapter were obtained by pooling the data from the S1, S2, and S3 national mathematics samples, with appropriate adjustments to the sampling weights (see tables 1.3 and 1.5 for student and school sample sizes, respectively) to provide a single set of "best" estimates for describing LEP in schools participating in the NAEP.

Results for the mathematics sample are presented in lieu of those for the science assessment since school and student sample sizes for the combined mathematics sample were typically slightly larger than those for the science sample. As the 1996 science assessment was conducted in the same schools as the mathematics assessments, results based on the science assessment samples differ little from those presented here.

General Background

Number of Years in the U.S. The SD/LEP questionnaire asked respondents to indicate how long each LEP student has lived in the United States. At each of the grades, a substantial percentage of the questionnaire respondents indicated that they did not know the answer to this question. This lack of knowledge on the part of the respondents complicates interpretation of the results, particularly at the eighth-grade, where 1 in 4 respondents did not know the answer to this question. The results for this question do suggest that twelfth-grade students were more likely to be recent arrivals to this country. Between 56 and 67[5] percent of twelfth-grade LEP students have lived in this country for five years or less while the corresponding range among fourth-grade LEP students is 37 and 48[6] percent. In contrast, between 42 and 53[7] percent of fourth-grade LEP students have lived in this country all their lives while the corresponding range among twelfth-grade LEP students is 10 to 21 percent[8]. A higher incidence of recent immigrants among grade 12 students seems to be a reasonable result. Many language minority students who immigrated to this country at a young age may have learned to speak and understand English to a degree that they are no longer classified as LEP.

Table 4.2 — Percentage distribution of time living in the U.S. for limited English proficient students by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

How long has this student lived in the U.S.?	Grade 4	Grade 8	Grade 12
All his/her life	42	16	10
More than 5 years	10	16	23
3-5 years	18	18	36
Less than 3 years	19	26	20
I don't know.	11	25	11

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

[5] Fifty-six percent of respondents indicated that the student lived in the U.S. for 5 years or less, while another 11 percent indicated they did not know the answer to the question. Since it is possible that some of the students in the "I don't know" category may have lived in the U.S. for less than five years, 67 percent would seem to be an upper bound for this percentage.

[6] Thirty-seven percent of respondents indicated that the student lived in the U.S. for 5 years or less, while another 11 percent indicated they did not know the answer to the question. Since it is possible that all of the students in the "I don't know" category may have lived in the U.S. all their lives, 48 percent would seem to be an upper bound for this percentage.

[7] Forty-two percent of respondents indicated that the student lived in the U.S. all his/her life, while another 11 percent indicated they did not know the answer to the question. Since it is possible that all of the students in the "I don't know" category may have lived in the U.S. all their lives, 53 percent would seem to be upper bound for this percentage.

[8] Ten percent of respondents indicated that the student lived in the U.S. all his/her life, while another 11 percent indicated they did not know the answer to the question. Since it is possible that some of the students in the "I don't know" category may have lived in the U.S. all their lives, 21 percent would seem to be an upper bound for this percentage.

Languages Spoken. As noted by August and Hakuta,[9] the largest proportion of LEP students are native speakers of Spanish, with the next largest groups comprised of speakers of one of a number of Asian languages. The SD/LEP questionnaire asked respondents to indicate each LEP student's first or native language. Response categories were "Spanish" and "Other." Respondents who chose "Other" were also provided with space to indicate the particular language spoken. Table 4.3 presents results for this question.

Table 4.3 — Percentage distribution of first or native language for limited English proficient students by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

What is student's first or native language?	Grade 4	Grade 8	Grade 12
Spanish	74	72	54
Other language	26	28	46

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Over 70 percent of LEP students at grades 4 and 8 were native speakers of Spanish. The magnitude of these percentages is quite close to the magnitude of those presented by August and Hakuta. There is somewhat greater diversity of language background among grade 12 LEP students. While native speakers of Spanish still make up the majority of LEP students, 46 percent are native speakers of some other language. The written responses of those choosing the "Other" option were coded into one of 17 distinct language categories. In general, "Other" responses were spread broadly across these categories at all three grades. The most frequently encountered other languages were Vietnamese, Hmong, Chinese, Russian, and Pacific-Island languages.

[9] August, D. & Hakuta, K. (Eds.). (1997) *Improving schooling for language-minority children: A research agenda.* Washington, DC: National Research Council. Institute of Medicine. Commission on Behavioral and Social Sciences and Education.

Regularity of School Attendance. The SD/LEP questionnaire contained two questions that were included to gather information about the recent schooling experiences of LEP students in the United States, or in the country from which they emigrated. The first of these questions asked how regularly the student has attended school in the United States or another country. The second question asked for the number of years that the student has been enrolled in a school where English was the primary language of instruction. Results on the regularity of school attendance are provided in table 4.4.

Table 4.4 — Percentage distribution of regularity of school attendance in the U.S. or another country for limited English proficient students by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Since reaching school age, how regularly has student attended school in the U.S. or another country?	Grade 4	Grade 8	Grade 12
Continuously	86	61	68
Intermittently	3	5	6
Little or not at all	2	2	2
I don't know.	9	32	23

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Respondents did not know the answer to this question for nine percent of fourth graders, 32 percent of eighth graders, and 23 percent of twelfth graders. Despite respondents' lack of knowledge in this area, the results suggest that the large majority of LEP students at all three grades have attended school regularly since reaching school age. At least 86 percent of fourth graders, 61 percent of eighth graders, and 68 percent of twelfth graders have attended school continuously in the U.S or another country. However, at least five percent, and as many as 14 percent, of fourth graders were reported to have attended school only intermittently, or not at all. At the eighth grade at least seven percent, and as many as 39 percent, have not have not attended school continuously. The corresponding percentages at the twelfth grade range from 8 to 31. Thus, in addition to learning English, it appears that some LEP students face additional academic and social challenges resulting from limited access to prior formal schooling, either here or in their native countries.

Enrollment in English-Language Schools[10]

Table 4.5 presents results on the number of years that students have been enrolled in schools where English was the primary language of instruction. This questionnaire item on schooling was included to differentiate between schooling a student may have received in his or her native country versus schooling in the U.S. This information sheds some light on the amount of students' exposure to the American school system and to American culture. At the fourth grade, respondents indicated that 14 percent of LEP students were attending schools where English was not the primary language of instruction. In contrast, among eighth- and twelfth-grade LEP students, nearly all were attending English language schools. As with some other previous questions, a substantial number of respondents did not know how long their students had been in English-language schools. This lack of knowledge on the part of respondents has methodological implications for NAEP. From 1990 through 1996, one of the NAEP-specified criteria for inclusion in the assessment was the number of years enrolled in a school where English is the primary language of instruction. The results from table 4.5 suggest that this information may be unavailable for a substantial percentage of LEP students.

Table 4.5 — Percentage distribution of number of years limited English proficient students have been enrolled in a school where English is the primary language of instruction by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Counting this year, how many years has this student been enrolled in a school where English is the primary language of instruction?	**Grade 4**	**Grade 8**	**Grade 12**
1 years	7	14	6
2 years	12	14	11
3-5 years	55	22	40
More than 5 years	4	30	32
English is not school's primary language of instruction	14	1	0
I don't know.	8	19	11

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

[10] The discussion in this section assumes that respondents knew whether the primary language of instruction in the school was English, and that those who indicated "I don't know" were not knowledgeable as to the duration of a student's enrollment in an English-language school.

The inclusion criteria in S1 indicated that LEP students could be excluded if they had been enrolled in a school where English is the primary language of instruction for less than 2 years. Those enrolled for 2 or more years were to be included. The results in table 4.5 suggest that, were these criteria strictly implemented, the vast majority of LEP students should have been included in NAEP. At least 85 percent of fourth-grade students, 67 percent of eighth-grade students, and 83 percent of twelfth-grade students have been enrolled in schools where English is the primary language of instruction for two or more years. Historically, NAEP inclusion rates for LEP students have typically been below the ideal minimums suggested by SD/LEP questionnaire results.

The Instructional Experiences of LEP Students

Current federal and state laws require schools and districts to provide special services to LEP students who are unable to participate meaningfully in an English-only school environment. A variety of such services is provided.[11,12] Many schools provide general or content-based instruction aimed at the development of English-language skills. Some provide academic instruction in English that is modified to be accessible to LEP students. Still others provide bilingual subject-matter instruction in which the students' native language is used to varying degrees. The 1996 SD/LEP questionnaire included a number of questions designed to collect information on the nature of the special services being received at that time by LEP students. Results based on these questions are presented in this section.

Special Instruction. Respondents were asked to indicate the number of years each student had been receiving specially designed academic instruction, including ESL, content-based ESL, sheltered English-content courses, native language support, and native language instruction. The results are presented in table 4.6.

The vast majority of LEP students at all three grades received some special instruction. Eighty-seven percent of grade 4 LEP students, 80 percent of grade 8 LEP students, and 81 percent of grade 12 students received special instruction. While many students had received such services for some time, relatively few students had been receiving special instruction for more than five years.

---

[11] August, D. & Hakuta, K. (Eds.). (1997) *Improving schooling for language-minority children: A research agenda.* Washington, DC: National Research Council. Institute of Medicine. Commission on Behavioral and Social Sciences and Education.

[12] Han, M., Baker, D., & Roderiguez, C. (1997). *A profile of policies and practices for limited English proficient students: Screening methods, program support, and teacher training.* (NCES Publication No. 97-472). Washington DC: National Center for Education Statistics.

Table 4.6 — Percentage distribution of years of enrollment in academic instruction specially designed for limited English proficient students by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Counting this year, how many years has this student been receiving academic instruction specially designed for students with limited English proficiency?	Grade 4	Grade 8	Grade 12
1 year	10	19	12
2 years	16	23	18
3-5 years	58	26	43
More than 5 years	3	13	7
Not receiving special instruction	13	20	19

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Academic Instruction in English.[13] As noted in chapter 1, the revisions to the NAEP inclusion criteria included a change in emphasis from time in English-language schools to number of years receiving academic instruction in English. The SD/LEP questionnaire asked how many years each student had been receiving academic instruction (mathematics, reading/ language arts, and science) primarily in English. Results are presented in table 4.7.

Despite the frequency of special instruction (see table 4.6), the large majority of LEP students at all three grades received academic instruction primarily in English. At grade 4, where native-language instruction was most frequently reported, only 1 in 4 LEP students was not receiving instruction primarily in English. Only six percent of eighth-grade LEP students and two percent of twelfth-grade LEP students were not receiving instruction primarily in English. As was the case for a number of other questions, substantial numbers of respondents were not able to answer this question. Given the role that this information plays in the revised NAEP inclusion criteria, these percentages are substantial and present a challenge to their consistent and rigorous implementation.

Table 4.7 — Percentage distribution of years of academic instruction in English for limited English proficient students by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Counting this year, how many years has this student been receiving academic instruction primarily in English?	Grade 4	Grade 8	Grade 12
1 year	12	16	8
2 years	10	15	14
3-5 years	41	19	39
More than 5 years	3	28	26
Not receiving instruction in English	26	6	2
I don't know.	8	17	10

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

[13] The discussion in this section assumes that respondents know whether a student is currently receiving instruction in a language other than English, and that those who indicated "I don't know' were not knowledgeable as to the duration of a students academic instruction in English.

Counting the current year, at least 44 percent of fourth-grade LEP students, 47 percent of eighth-grade LEP students, and 65 percent of twelfth-grade LEP students have received academic instruction primarily in English for 3 or more years. Since the revised inclusion criteria state that all such students should be included in NAEP, these percentages represent the minimum inclusion rates that should be expected with the revised criteria. As discussed in chapter 5, where the revised criteria were used, the percentages of LEP students that were assessed exceeded these minimums in all but one instance.

Instructional Practices. The SD/LEP questionnaire also asked about instructional practices in three specific content areas: reading/language arts, mathematics, and science. For each content area, respondents were asked to indicate which of three options applied to a given student: 1) specially designed instruction in English (such as ESL), 2) native language instruction, or 3) mainstreamed with no specially designed instruction. For purposes of analysis, the responses by content area (questions 46, 47 and 48 on the questionnaire) were combined with responses from the question on special instruction reported in table 4.6 (question 40 on the questionnaire). Specifically, students were classified as receiving no special instruction in a specific content area if 1) they were so classified on the basis of question 46, 47, or 48; or 2) if question 46, 47, or 48 was omitted, but the student was classified on the basis of question 40 as receiving no special instruction (see questionnaire in appendix B).

Results for each of the three content areas (reading/language arts, mathematics, and science) are presented in table 4.8. A number of striking and, compared to table 4.6, contradictory results appear. First, the subject area results (table 4.8) seem to suggest that fewer students received special instruction than was indicated by the general question on academic instruction (table 4.6). For example, from table 4.8, the percentages of LEP students in grade 4 reported to be receiving some special instruction were 79 for reading/language arts, 60 for mathematics, and 61 for science. The corresponding percentage from table 4.6 was 87. Similarly, the percentages reported at grades 8 and 12 on table 4.8 all exceed the percentages reported for grades 8 and 12 in table 4.6. Second, the results in table 4.6 suggest little differences by grade in the percentages of students receiving special services. In contrast, the results in table 4.8 suggest that fewer students appeared to receive special instruction at the higher grades. For example, the percentages of LEP students who received special instruction in reading/language arts were 79 percent at grade 4, 65 percent at grade 8, and 60 percent at grade 12. This same pattern of apparent decrease over grades is evident in the responses for the mathematics and science results.

Table 4.8 — Percentage of limited English proficient students receiving specially designed instruction in selected content areas by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Instructional Area	Grade 4	Grade 8	Grade 12
Reading/language Arts			
Special instruction in English	57	59	60
Native language instruction	22	6	0
No special instruction	21	35	40
Mathematics			
Special instruction in English	34	32	19
Native language instruction	27	8	8
No special instruction	40	60	73
Science			
Special instruction in English	35	36	22
Native language instruction	26	6	6
No special instruction	39	58	72

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

In the absence of follow-up research, such as interviews with questionnaire respondents, one can only conjecture as to why such differences in results based on these two sets of questions could have arisen. Part of the difference is no doubt due to the fact that question 40 asked generally about academic subjects, while questions 46 through 48 asked about specific subject areas. Question 40 did not specifically define the term "academic areas" and respondents may have interpreted this term quite broadly, resulting in their classifying more students as receiving special instruction. In particular, respondents may have classified students in ESL classes quite differently in question 40 than in questions 46 through 48.

August and Hakuta[14] (1997) distinguish between two types of instructional approaches for teaching English: ESL, and content-based ESL. They define ESL as instruction aimed at the development of English-language skills, with a primary focus on grammar, vocabulary, and communication rather than academic content area. Content-based ESL is defined by August and Hakuta as periods of ESL instruction that are structured around academic content, rather than generic language skills. August and Hakuta also note that these two approaches are sometimes, but not always, combined with subject-matter instruction delivered in one of a number of program models, such as sheltered instruction, structured immersion, or transitional

[14] August, D. & Hakuta, K. (Eds.). (1997). *Improving schooling for language-minority children: A research agenda.* Washington, DC: National Research Council. Institute of Medicine. Commission on Behavioral and Social Sciences and Education.

bilingual education. Thus, some LEP students may be receiving general or content-based ESL instruction but no special subject-matter instruction, while other students may be receiving ESL instruction along with special subject-matter instruction. It is possible that respondents classified some or all of the students in the first group (i.e., ESL instruction without special subject-matter instruction) as receiving special instruction when asked the general question about academic instruction. However, when asked questions 46 to 48 in the context of specific subject areas, respondents may have classified these same students as receiving no specially designed instruction.

The results in table 4.8 are reasonably consistent with the results presented earlier in indicating that, particularly at the higher grades, relatively few students received native language instruction in academic areas. At grade 4, the percentages of LEP students who received native language instruction in reading/language arts, mathematics, and science were 22, 27, and 26 percent, respectively. At the higher grades, the percentages range from 0 percent (grade 12 for reading/language arts) to 8 percent (grades 8 and 12 for mathematics). It is interesting to note that the percentages of LEP students receiving native language instruction in these areas are slightly higher than the results reported in table 4.7. Discrepancies are most likely due to question wording. The question on which the results in table 4.7 are based sought the number of years that the student was receiving academic instruction *primarily* in English.

Grade-Level of English-Language Instruction. As noted in August and Hakuta (1997)[15] the specific nature of English language instruction for LEP students may vary. Some LEP students are fully mainstreamed and receive the same instruction as English-proficient students, while others (e.g., students in sheltered instruction, in structured immersion programs, or in transitional bilingual programs) may be receiving subject-matter instruction modified to be accessible to them. The SD/LEP questionnaire included a question asking what grade level of instruction in English the student was receiving in reading/language arts, mathematics, and science.

Results on the grade level of English-language instruction are presented in table 4.9. Among LEP students receiving English-language instruction, the majority received instruction at grade level in all grades and subjects. At grades 4 and 8, greater percentages of students received English-language instruction in math and science than in reading/language arts. Grade 12 appears more uniform in that 62 to 64 percent received grade-level instruction across the three subject areas.

[15] Ibid.

Table 4.9 — Percentage distribution of limited English proficient students by grade level of English language instruction in reading/language arts, mathematics, and science by grade: NAEP 1996 mathematics sample

What grade level of instruction in the English language is this student currently receiving in:	Grade 4	Grade 8	Grade 12
Reading/language arts			
Above grade level	0	0	0
At grade level	58	62	62
One year below grade level	24	7	8
Two or more years below grade level	17	31	30
Mathematics			
Above grade level	1	1	0
At grade level	83	74	64
One year below grade level	11	7	10
Two or more years below grade level	5	18	26
Science			
Above grade level	0	0	0
At grade level	83	76	64
One year below grade level	11	5	10
Two or more years below grade level	6	19	26

NOTE: Percentages are based on students receiving English-language instruction. Students receiving native-language instruction are not included in the results. Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

In chapter 3 it was noted that, at the higher grades, students with disabilities appear to be further behind their nondisabled peers with respect to grade-level instruction. A similar pattern is evident among LEP students receiving instruction in English. For example, the percentage who received mathematics instruction two or more years below grade level was 5 percent at grade 4 compared with 26 percent at grade 12. Similar results hold for science. In reading/language arts, the percentage of students two or more years below grade level was 17 percent at grade 4 compared with 30 percent at grade 12. It should be noted, however, that NAEP is a cross-sectional survey, which makes the interpretation of such cross-grade patterns somewhat complicated. Changes in percentages of students who received grade-level instruction are confounded with differences in the demographic makeup of the grade cohorts and dropout rates at the higher grades.

Academic Performance Levels. As noted earlier, many LEP students come from disadvantaged socioeconomic backgrounds and are clustered in large schools located in traditionally low-achieving urban areas. It is not surprising then that the present data suggest low academic performance for LEP students in comparison to their English-proficient peers. According to Han *et al*.,[16] LEP students receive lower grades, score below their classmates on standardized reading and mathematics tests, and are often judged by their teachers as academic "underachievers." NAEP results from 1990 through and including 1996 are consistent in showing average scores that are lower for the LEP students included in the NAEP assessment than for their English-proficient peers in geography, reading, mathematics, science, writing, and U.S. history.[17]

The SD/LEP questionnaire included a question that asked respondents to indicate current performance levels in English in reading/language arts, mathematics, and science. As with a number of similar questions, substantial percentages of respondents indicated that they did not know the students' current grade level of performance in English. However, results based on those that did reply suggest that among LEP students receiving English language instruction, a significant number were performing below grade level in English. In reading/language arts, where one might expect the impact of limited language proficiency to be most pronounced, higher percentages of students were judged as performing below grade level than in math or science.

---

[16] Han, M., Baker, D., & Roderiguez, C. (1997). *A profile of policies and practices for limited English proficient students: Screening methods, program support, and teacher training.* (NCES Publication No. 97-472). Washington DC: National Center for Education Statistics.

[17] Results for each of the subject areas are available on the World Wide Web at the NCES website (http://nces.ed.gov/nationsreportcard/site/home.asp) in the Student Data sections of the Summary Data Tables.

Table 4.10 — Percentage distribution of estimated grade level of performance in reading/language arts, mathematics, and science for limited English proficient students by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

At what grade level is this student currently performing in the English Language in:	Grade 4	Grade 8	Grade 12
Reading/language arts			
Above grade level	1	2	8
At grade level	26	24	29
One year below grade level	41	17	14
Two or more years below grade level	29	45	36
I don't know.	3	13	13
Mathematics			
Above grade level	2	2	8
At grade level	52	35	33
One year below grade level	27	16	11
Two or more years below grade level	14	30	22
I don't know.	5	17	26
Science			
Above grade level	1	1	7
At grade level	46	31	32
One year below grade level	27	13	10
Two or more years below grade level	20	34	20
I don't know.	6	22	31

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Testing Accommodations and Adaptations

As discussed in chapter 1, the NAEP program is introducing changes in its inclusion criteria and its policies on the provision of accommodations and adaptations that were designed to increase the meaningful participation of LEP students in NAEP. Districts and states are, in many cases, also involved in such activities. NAEP program policy is built around the concept of testing students in NAEP in the manner most similar to that used for district and statewide achievement testing. In light of this, the SD/LEP questionnaire included two questions pertaining to the typical use of accommodations and adaptations in achievement testing with each student.

Respondents were asked to indicate whether any accommodations or adaptations are used for achievement testing with each student. Results for this question are presented in table 4.11. The results suggest that accommodations or adaptations are not routinely provided. Respondents indicated that about one-third or less of LEP students use accommodations and adaptations in achievement testing. In the S3 samples of the 1996 NAEP mathematics assessment, 30 percent of grade four LEP students, 18 percent of grade eight LEP students, and 6 percent of grade twelve LEP students were assessed with an accommodation or adaptation.

Table 4.11 — Percentage distribution for limited English proficient students of whether accommodations or adaptations are used for achievement testing by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Are any accommodations or adaptations used for achievement testing for this student?	Grade 4	Grade 8	Grade 12
Yes	36	27	22
No	64	73	78

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Respondents were also asked to indicate which of a number of accommodations or adaptations are used for each student. Results for this question are presented in table 4.12. Of the seven varieties of accommodations provided, no single variety is predominant. The highest percentage (22 percent) was the use of native language versions of tests in grade 4. However, this usage appeared to drop off quickly in grades 8 and 12. None of the six other forms of accommodations exceeded 12 percent usage in any grade.

Table 4.12 — Percentage of limited English proficient students receiving selected accommodations/adaptations in achievement testing by grade: NAEP 1996 mathematics sample

Accommodation/adaptation	Grade 4	Grade 8	Grade 12
Native language version of test	22	8	3
English/native language dictionary	4	4	11
Word lists or glossaries	6	2	4
Extended time	10	12	10
Help with directories and questions	11	12	6
Direction read aloud twice in English	11	11	6
Questions read aloud twice in English	11	9	6
Other	3	4	4

NOTE: Multiple responses were permitted.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Chapter Summary

This chapter provided an overview of the demographic characteristics and instructional experiences of the 1996 NAEP population of LEP students, as reflected in the questionnaire filled out for each LEP student by a knowledgeable school staff member. The results presented are based on the 1996 national NAEP mathematics sample.

LEP students at the higher grades are more likely to be recent immigrants. At least 56 percent of twelfth-grade LEP students have lived in this country for five years or less. Fewer fourth-grade LEP students, no more than 48 percent, have lived in this country for five years or less. At least 42 percent of the fourth-grade LEP students have lived in the United States all their lives. The percentages of twelfth-grade students who have lived in this country all their lives is considerably lower (no more than 21 percent).

The largest proportion of LEP students speaks Spanish as their native language. The most frequently encountered languages other than Spanish were Vietnamese, Hmong, Chinese, Russian, and Pacific-Island languages.

The vast majority of LEP students have attended school regularly since reaching school age. However, at least 5 percent of fourth-graders, 7 percent of eighth graders, and 8 percent of twelfth graders were reported to have attended school only intermittently or not at all.

From 1990 through 1992, one of the NAEP criteria for including LEP students involved number of years enrolled in English-language schools. Most LEP students (at least 85 percent at grade 4, 67 percent at grade 8, and 83 percent at grade 12) had been enrolled in a school for two or more years where English was the primary language of instruction. Among students at grades 8 and 12, almost none attended schools where English is not the primary language of instruction. However, at grade 4, respondents indicated that 14 percent of LEP students attended schools where English is not the primary language of instruction. As with several other questions, the percentages of missing data were fairly high. The fact that these data may not be known for substantial percentages of examinees clearly raises questions as to its appropriateness and utility in making inclusion decisions.

The vast majority of LEP students at all three grades received some special instruction. Eighty-seven percent of grade four LEP students, 80 percent of grade eight LEP students, and 81 percent of grade twelve students received special instruction. While many students have received such services for some time, relatively few students at grades 8 and 12 have received special instruction for more than five years.

At least forty-four percent of grade four LEP students, 47 percent of grade eight LEP students, and 65 percent of grade twelve LEP students received academic instruction primarily in English for 3 or more years. Since the revised inclusion criteria state that all such students should be included in NAEP, these percentages represent the minimum inclusion rates that should be expected with the revised criteria. These expected minimums were generally exceeded by the actual S2 and S3 inclusion percentages in the 1996 NAEP assessments.

At grades four and eight, higher percentages of LEP students received instruction at their grade level in mathematics and in science (about 80 percent at grade 4 and 75 percent at grade 8) than in reading/language arts (about 60 percent at both grades).

Consistent with other known performance data on LEP students, NAEP questionnaire results suggest low levels of academic performance for most of these students. Regarding academic performance in reading/language arts, where one might expect the impact of limited-language proficiency to be most pronounced, at least half of LEP students at all three grades were judged as performing below grade level in English. In mathematics and science, the percentages reported performing below grade-level range from at least 30 percent in science to at least 47 percent.

The questionnaire results suggest that accommodations or adaptations are not routinely provided for LEP students. Respondents indicated that about one third or less of LEP students use accommodations and adaptations in achievement testing. The usage of seven types of accommodations was fairly evenly split for students with limited English proficiency. The most common accommodation was the use of the Spanish language mathematics booklet in grade 4. Of the six other types of accommodations offered, none exceeded 12 percent usage.

Chapter 5

The Inclusion of Limited English Proficient Students in NAEP: A Closer Look

As discussed in chapter 1 of this report, the 1996 NAEP assessments were redesigned to allow the introduction and evaluation of modified procedures intended to increase participation in NAEP among limited English proficient students. The intent of the design change was to assess these students in ways that allow them to best demonstrate their subject-matter knowledge. Modifications were made in two specific areas. First, based on the recommendations of an advisory committee, changes were made to the specific instructional variables on which inclusion decisions were to be based. Second, assessment accommodations and adaptations were offered to LEP students who are regularly tested with them. In particular, a Spanish/English bilingual version of one form of the NAEP mathematics assessment was made available at grades 4 and 8. The three subsamples used in the 1996 national NAEP design were described in Chapter 1.

Initial evaluations of the effect of these modifications on inclusion rates for LEP students were presented in the *NAEP 1996 Mathematics Report Card for the Nation and the States* and the *NAEP 1996 Science Report Card for the Nation and the States*.[1] The results indicated the following:

- The introduction of revised inclusion criteria, without the provision of accommodations, had no clear effect on the percentage of LEP students who were assessed in NAEP. Absent the provisions of accommodations, at grades 4 and 12, national inclusion rates for the mathematics assessment appeared lower where revised criteria were used (S2) than where the original criteria for inclusion were used (S1). However, these differences were not statistically significant. At grade 8, national inclusion percentages for these samples were quite comparable. State NAEP inclusion rates at grades 4 and 8 showed no clear pattern.

---

[1] Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). *NAEP 1996 mathematics report card for the nation and the states*. Washington, DC: National Center for Education Statistics.
O'Sullivan, C.Y., Reese, C.M., & Mazzeo, J. (1997). *NAEP 1996 science report card for the nations and states.* Washington, DC: National Center for Education Statistics.

- For the mathematics assessment, where accommodations were permitted and a Spanish/English bilingual version of the test was made available at grades 4 and 8, the provision of accommodations and adaptations did increase the percentages of LEP students that were assessed at grades 4 and 8, but not at grade 12. At grades 4 and 8, the percentage of LEP students included in NAEP was higher where accommodations and adaptations were offered (S3) than where accommodations and adaptations were not offered (S2). Both these samples used the revised inclusion criteria. The inclusion percentages at these grades also appeared higher in the accommodated sample than in the sample where accommodations and adaptations were not offered, but the original inclusion criteria were used (S1). However, these latter differences were not statistically significant. None of the grade-12 differences was statistically significant.
- For the science assessment, where accommodations were permitted but no bilingual version of the test was made available, the provision of accommodations had no clear effect on LEP inclusion percentages. At grades 4 and 12, inclusion percentages appeared higher in S3 than in S2. However, these apparent differences were not statistically significant. At grade 8, S2 and S3 inclusion percentages were comparable.

This chapter looks more closely at NAEP inclusion rates among LEP students from the mathematics assessment. In the first section of the chapter, results originally included in the *NAEP 1996 Mathematics Report Card* are presented again and discussed in light of additional information available from the SD/LEP questionnaire. In the second section of this chapter, inclusion rates among particular segments of the LEP population are discussed. An additional purpose of the latter section is to provide a more detailed understanding of the impacts of the revised inclusion criteria and the provision of accommodations among these particular segments. The focus is placed on those segments that one would have expected to be particularly affected by the changes in wording made to inclusion criteria or to the provision of accommodations.

Assessing the Impacts of Criteria Revisions and the Provision of Accommodations and Adaptations

The results presented in chapter 4 suggest that a large percentage of LEP students are receiving instruction in English that is at grade level. Based on these results, one might expect high inclusion rates among LEP students where appropriate accommodations and adaptations can be made available.

Table 5.1 presents the percentages of LEP students who were assessed under standard conditions (i.e., in English without accommodations or adaptations), assessed with accommodations or adaptations (including the Spanish/English bilingual test booklet), or excluded from the assessment for each of the three national samples. These data were originally included in a slightly different format in the *NAEP 1996 Mathematics Report Card.*[2]

Table 5.1 — Percentage of limited English proficient students included in NAEP, by sample type and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

	Assessed			Excluded
	Without accommodations	With accommodations	Total	**Total**
Grade 4				
S1[1]	61	—	61	39
S2[2]	41‡	—	41	59‡
S3[3]	47	30	76††	24††
Grade 8				
S1[1]	60	—	60	40
S2[2]	63	—	63	37
S3[3]	61	18	78††	22††
Grade 12				
S1[1]	84	—	84	16
S2[2]	73	—	73	27
S3[3]	81	6	87	13

[1] 1990-1996 operational inclusion critera

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

‡ Indicates a significant difference between S1 and S2 results.

†† Indicates a significant difference between S2 and S3 results.

— Not applicable because accommodations were not offered.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

[2] The difference in S1 and S2 inclusion rates at grade 4 was not reported as statistically significant in the *1996 Mathematics Report Card.* Because of the research and development nature of this report, significance tests were not adjusted for multiple comparisons except in chapter 7, where large numbers of comparisons were made. The 1996 report cards used multiple comparison procedures that assumed a family size of 3 (the number of grades) for each comparison of samples.

For the samples in which accommodations and adaptations were not offered (i.e., S1 and S2), 60 percent or more of LEP students were included with one exception (the S2 sample at grade 4). Where accommodations and adaptations were available (i.e., in the S3 samples), the total percentages of assessed LEP students ranged from 76 percent (at grade 4) to 87 percent (at grade 12). The percentage of LEP students actually assessed with accommodations or adaptations appeared smallest at grade 12, where the Spanish/English bilingual version was not provided.

The results of two questions from the NAEP SD/LEP questionnaire provide an additional context within which to discuss NAEP inclusion rates, with and without the provision of accommodations. Questionnaire respondents were asked whether or not students could meaningfully participate in NAEP without accommodations or adaptations. Table 5.2 presents results based on this question for the combined mathematics sample. It is of particular interest to compare the percentages in table 5.2 with those in table 5.1

Table 5.2 — Percentage of limited English proficient students who could meaningfully participate in NAEP without accommodations or adaptations by grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD

In your judgment, could this student meaningfully participate in NAEP without accommodations?	Grade 4		Grade 8		Grade 12	
	Yes	No	Yes	No	Yes	No
Mathematics	46	54	60	40	72	28

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

With one exception (the S2 sample at grade 4), the percentages of LEP students actually assessed in NAEP without accommodation or adaptations (table 5.1) match or exceed the estimates indicated by the questionnaire respondents (table 5.2). These results suggest that increases in the numbers of LEP students are not likely to result solely from revisions to inclusion criteria that do not also involve the provision of accommodations and adaptations.

A second SD/LEP questionnaire item provides a somewhat different perspective from which to evaluate current levels of inclusion and the degree of improvement that might be attainable. Questionnaire recipients were asked to indicate the conditions under which each student would participate in NAEP if accommodations and adaptations were made available: 1) in English without accommodations or adaptations; 2) in English with the accommodations or adaptations specified for achievement testing with the student; 3) in his or her native language without accommodations or adaptations; 4) in his or her native language with the accommodations or adaptations specified for achievement testing with the student; or 5) the student would not participate. Results based on the combined sample from the mathematics assessment are presented in table 5.3.

Table 5.3 — Percentage distribution of limited English proficient students who would participate in NAEP by accommodation status and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

How would this student participate in NAEP?	Grade 4	Grade 8	Grade 12
In English			
Without accommodation/adaptation	25	46	55
With accommodation/adaptation	27	22	30
Total	51	69	86
In native language			
Without accommodation/adaptation	28	16	8
With accommodation/adaptation	9	6	1
Total	37	22	9
Student would not participate	11	9	6

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Questionnaire responses indicate that 11 percent or less of LEP students would not participate in NAEP. Actual exclusion rates in NAEP for LEP students (table 5.1) were higher than this, however, even when accommodations and adaptations were offered. Thus, these results suggest that further modest improvements in inclusion might still be possible.

Further corroboration for this conjecture can be seen by examining the data on the percentage of students that could be tested in English without accommodations and adaptations. The percentages indicated in table 5.3 (25 percent at grade four, 46 percent at grade eight, and 55 percent at grade twelve) all appear lower than the percentages actually assessed without accommodations or adaptations in the S3 samples (47 percent at grade four, 61 percent at grade eight, and 81 percent at grade twelve). Thus, some of the LEP students tested in NAEP under standard conditions may have been tested with accommodations or adaptations if the appropriate types had been offered.

Other results in table 5.3 suggest that expanding the amount of permitted native-language testing (e.g., by providing alternate forms in languages other than Spanish, or by allowing oral administration of the test in the student's native language) would be the most likely means of expanding inclusion. About as many LEP students were actually assessed in English as respondents reported would be assessed in English. According to the questionnaire responses, 51 percent of grade-four LEP students, 69 percent of grade-eight LEP students, and 86 percent of grade-twelve LEP students would have participated in NAEP in English. These percentages are quite close to the S3 inclusion percentages for the 1996 NAEP science assessment (53 percent, 64 percent, and 83 percent for grades 4, 8 and 12, respectively), as reported in the *NAEP 1996 Science Report Card*.[3] In the 1996 NAEP science assessment, accommodations and adaptations (including use of a Spanish/English bilingual glossary and bilingual dictionaries) were permitted in the S3 sample, but no alternate-language testing was permitted. Thus, it appears that the English-language accommodations and adaptations in the 1996 NAEP science assessment were sufficient to include nearly all LEP students capable of meaningful testing in English. Additional improvements in inclusion would seem to depend on the expansion of native-language testing. Questionnaire responses indicate that 37 percent of grade-four LEP students, 22 percent of grade-eight LEP students, and 9 percent of grade-twelve LEP students would have participated in NAEP in their native languages.

For reasons of cost and feasibility, NAEP could not provide all possible accommodations and adaptations in all grades and subjects. Nor did the program offer all LEP students the opportunity to test with accommodations or adaptations. Accommodations or adaptations were to be provided in NAEP only for students who routinely receive them in classroom, school, or district testing. For the most part, NAEP permitted any and all accommodations usually provided by the school that would not interfere with the intent of the assessment. The list of permitted accommodations included the use of bilingual dictionaries. The adaptations offered included a Spanish/English bilingual booklet at grades 4 and 8 in mathematics.

[3] O'Sullivan, C.Y., Reese, C.M., & Mazzeo, J. (1997). *NAEP Science Report Card for the nation and the states*. Washington, DC: National Center for Education Statistics.

Table 5.4 presents results from the mathematics assessment on the relative frequency of use for the various accommodations and adaptations among S3 LEP students.

Table 5.4 — Percentage of limited English proficient students assessed with each of the offered accommodations/adaptions by grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

Accommodation/adaptation	Grade 4	Grade 8	Grade 12
Bilingual test booklet	24	14	*
One-on-one testing	1	0	2
Extended time (regular session)	2	0	2
Small group session	2	0	0
Directions read aloud (regular session)	0	2	0
Bilingual dictionary	0	1	0
Other	0	0	1

NOTE: Bilingual test booklet was not offered at grade 12.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Where it was available (i.e., at grades 4 and 8), the principal nonstandard testing condition used was the Spanish/English bilingual test booklet. At grade 12, where the booklet was not available, one-on-one testing and extended time were the most commonly used nonstandard administrations. However only four percent of grade 12 LEP students were tested in such sessions.

A Look at Inclusion Rates within Selected Subgroups

As noted previously, changes were made to NAEP inclusion criteria for LEP students that were intended to provide a better match between their English literacy levels and the manner in which they would participate in NAEP. One part of this change involved a switch from using the number of years that the student attended a school where English was the primary language of instruction to using the number of years that the student had been receiving academic instruction in English. The original criteria, which were developed for a NAEP in which accommodations and adaptations were not offered, stated that LEP students could be excluded if they had been enrolled for less than two years in schools where English was the primary language of instruction and if they were unable to meaningfully participate in the assessment. Thus, strict application of these criteria would require that all students enrolled in English-language schools for two or more years be included in NAEP. The revised criteria, which were designed for a NAEP in which accommodations and adaptations were to be offered, stated that LEP students should be included unless they had received academic instruction in English for less than three years and were unable to meaningfully participate in NAEP. Thus, strict application of these criteria would require all students receiving academic instruction in English for three or more years to be assessed.

A close review of these criteria suggests the possibility that, absent the provision of accommodations and adaptations, the revised criteria may, in fact, be less inclusive than the original criteria. It seems reasonable to assume that most, if not all, students enrolled in English-language schools for less than two years have also received academic instruction in English for less than three years. However, consider LEP students who have attended English-language schools for two years or more. Among this group, some LEP students may have been receiving academic instruction in English for less than three years and some for three years or more. It was hoped that the provision of accommodations and adaptations would enable both groups of students to participate at high levels. However strict application of the criteria in the absence of accommodations and adaptations could result in students in the first group (i.e., in English-language schools for two or more years, but receiving academic instruction in English for less than three years) being included less frequently under the revised criteria than under the original criteria.

Enrollment in English-Language Schools

The SD/LEP questionnaire contained a question designed to allow some evaluation of whether the revised criteria were resulting in less rather than more inclusion for LEP students. As noted in chapter 4, each respondent was asked to indicate for each student the number of years of enrollment in a school where English was the primary language of instruction. Based on the responses, students were split into two groups, 1) those enrolled for less than two years, and 2) those enrolled for two or more years. For students in the first group (i.e., LEP students enrolled for less than two years) one would expect to see little difference in inclusion rates because exclusion was likely to occur under both criteria. However, for the second group (i.e., LEP students enrolled for two or more years), one might expect to see lower inclusion rates with the revised criteria than with the original criteria, particularly in the absence of accommodations and adaptations.

Inclusion rate results by years of enrollment in English-language schools are presented in table 5.5. The results are equivocal. There is some evidence to support the conjecture that, absent accommodations and adaptations, inclusion rates would be lower using the revised criteria. Among students enrolled in English-language schools for two or more years, inclusion rates appeared lower in S2 than in S1 at all three grades. However, only at grade 4 was the apparent difference statistically significant. Somewhat surprisingly, inclusion rates were also lower in S2 than in S1 for fourth-grade LEP students enrolled less than two years. Such a result seems anomalous because exclusion of these students is permitted under both sets of criteria and raises some concerns as to reasons for the differences in S1 and S2 inclusion rates.

The results in table 5.5 also suggest that under the revised criteria the provision of accommodations and adaptations will increase participation rates for both groups of students. The evidence for this increase is strongest at grade 4, where a greater percentage of LEP students was included in S3 than in S2, regardless of years enrolled in English-language schools. The results at grade 8 and grade 12 follow a similar pattern, with inclusion appearing higher in S3 than in S2 in all cases. However, none of the grade 8 and grade 12 differences are statistically significant. Table 5.5 results further suggest that using the revised criteria in conjunction with the offering of accommodations did not increase inclusion except in one instance. At grade 4, inclusion rates were higher in S3 than in S1 among students enrolled in English-language schools for less than two years. However, in all other instances where sample sizes were adequate, S3 and S1 inclusion rates appeared similar and differences were not statistically significant.

Table 5.5 — Percentages of limited English proficient students assessed by years enrolled in English-language school, by sample type and grade: 1996 NAEP mathematics sample

Counting this year, how many years has this student been enrolled in a school where English is the primary language of instruction?	Less than 2 years			2 years or more		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4						
S1[1]	26	44	56	82	78	22
S2[2]	63	16 ‡	84 ‡	161	54 ‡	46 ‡
S3[3]	35	92 † ††	8 † ††	170	72	28
Grade 8						
S1[1]	15	– –	– –	51	80	20
S2[2]	26	16	84	150	68	32
S3[3]	24	44	56	73	81	19
Grade 12						
S1[1]	1	– –	– –	30	84	16
S2[2]	11	– –	– –	147	77	23
S3[3]	10	– –	– –	107	89	11

[1] 1990-1996 operational inclusion critera

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

– – Indicates insufficient sample size to permit a reliable estimate.

‡ Indicates a significant difference between S1 and S2 results.

† Indicates a significant difference between S1 and S3 results.

†† Indicates a significant difference between S2 and S3 results.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics *Assessment.*

Years Receiving Academic Instruction in English

The revised inclusion criteria stated that all students receiving academic instruction in English for three or more years were to be included in NAEP. Thus, one should expect 100 percent inclusion of students in this category in the S2 and S3 samples. Moreover, since it is reasonable to expect that virtually all LEP students receiving academic instruction in English for three or more years have also been enrolled in English-language schools for two or more years, high inclusion rates should also be expected in the S1 sample. The SD/LEP questionnaire contained a question that allows some evaluation of the degree to which these ideal rates were achieved. Respondents were asked to indicate the number of years each LEP student had been receiving academic instruction in English. Two groups of students were formed based on the responses to this question: 1) LEP students with less than 3 years academic instruction in English; and 2) LEP students with 3 or more years instruction. Inclusion results for these two groups of students are shown in table 5.6.

Table 5.6 — Percentages of limited English proficient students assessed by years receiving academic instruction in English, by sample type and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

Counting this year, how many years has this student been receiving academic instruction primarily in English?	Less than 3 years			3 years or more		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4						
S1[1]	66	61	39	36	85	15
S2[2]	124	19 ‡	81 ‡	107	67 ‡	33 ‡
S3[3]	96	68 ††	32 ††	108	85 ††	15 ††
Grade 8						
S1[1]	35	41	59	26	89	11
S2[2]	71	41	59	100	77	23
S3[3]	58	62	38	51	85	15
Grade 12						
S1[1]	5	– –	– –	26	92	8
S2[2]	46	71	29	112	76	24
S3[3]	30	78	22	89	92	8

[1] 1990-1996 operational inclusion critera

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

– – Indicates insufficient sample size to permit a reliable estimate.

‡ Indicates a significant difference between S1 and S2 results.

†† Indicates a significant difference between S2 and S3 results.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

In general, inclusion rates appear higher among students with three or more years of academic instruction in English than for students with less than three years. This pattern of results was evident, with one or two exceptions, over all grades and samples. However, it is also evident that not all students reported to be receiving academic instruction in English for more than three years were included in the assessment. The percentages of such students included in NAEP range from 92 (at grade 12 in the S1 and S3 samples) to 67 (at grade 4 in the S2 sample). These less-than-perfect levels of inclusion could be the result of a number of factors, including incorrect application of the inclusion criteria and invalid or incorrect questionnaire responses.

There are additional aspects of the table 5.6 results that should be noted. Once again, in the absence of accommodations and adaptations, grade-four inclusion rates were significantly lower with the revised criteria (i.e., in S2) than with the original criteria (i.e., in S1). Such a result seems plausible for the "less-than-three-years" group, since some of these students may have been enrolled in English-language schools for more than two years, and hence included on the basis of the original criteria. However, S2 inclusion rates were also lower than S1 rates for the "three-years-or-more" group. Since this group would appear to meet both the original and revised criteria, the reasons for this difference remain unclear.

The results in table 5.6 also suggest that when the revised criteria are used, the provision of accommodations and adaptations resulted in increased inclusion at grade 4 for both groups of students. Inclusion rates in S3 were significantly higher than those in S2 for both groups of students. Inclusion rates at grades 8 and 12 also appeared higher for S3 than those in S2 for both groups of students. However, these apparent differences were not statistically significant.

However, the joint effect of changing inclusion criteria and providing accommodations had no measurable effect on inclusion rates for either of the two groups. Differences between S3 and S1 inclusion rates at all three grades followed no consistent pattern and were not statistically significant.

Inclusion Rates by Presence/Absence of Accommodations

As noted earlier, not all LEP students in the S3 sample were offered the opportunity to test with accommodations or adaptations. Accommodations or adaptations were to be provided in NAEP only to those students who regularly received them in school, district, or statewide testing. As part of the SD/LEP questionnaire, respondents were asked to indicate for each LEP student whether accommodations or adaptations are regularly used for achievement testing. Two categories of response were permitted: yes or no. Table 5.7 presents results on NAEP inclusion percentages for these two responses. Results are presented separately for the S1, S2, and S3 samples at all three grades.

Table 5.7 — Percentages of limited English proficient students assessed by whether accommodations or adaptations are normally used, by sample type and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

Are any accommodation/ adaptations used for achievement testing for this student?	YES				NO			
	N	Assessed standard	Assessed with accommodations	excluded	N	Assessed Standard	Assessed with Accommodations	Excluded
Grade 4								
S1[1]	31	52	—	48	95	68	—	32
S2[2]	81	24	—	76 †	169	54	—	46
S3[3]	109	31	39	30 ††	103	75	10	15 ††
Grade 8								
S1[1]	19	– –	—	– –	60	82	—	18
S2[2]	63	39	—	61	137	71	—	29
S3[3]	47	33	35	32 ††	114	79	6	15
Grade 12								
S1[1]	2	– –	—	– –	31	84	—	16
S2[2]	30	67	—	33	138	78	—	22
S3[3]	30	54	16	30	102	88	5	7

[1] 1990-1996 operational inclusion critera

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

– – Indicates insufficient sample size to permit a reliable estimate.

— Not applicable because accommodations were not offered.

† Indicates a significant difference between S1 and S3 results.

†† Indicates a significant difference between S2 and S3 results.

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

For students who normally receive accommodations, inclusion rates increased for grade four and grade eight groups when accommodations were made available. Inclusion rates in S3 were higher than those in S2 for both grades. At grade 4, inclusion rates also appeared in S3 than in S2. However, this difference was not statistically significant. Some LEP students who do not usually receive accommodation in testing were apparently provided accommodations in the NAEP assessment. The percentages in this category were small (10, 6, and 5 percent in grades 4, 8, and 12, respectively). Again, this result raises methodological issues about application of procedures in the field, or the validity of questionnaire responses.

Inclusion Rates by Testing in English Versus Other Language

As noted in August and McArthur,[4] great diversity exists among the LEP school population regarding their home and educational language background. Regardless of language background, some LEP students are newly arrived immigrants with high literacy skills and good school experiences in their native languages, while other LEP students are newly arrived immigrants with low literacy skills and limited school experiences in their native language. Still others have been schooled exclusively in the United States and instructed in bilingual classes or predominantly English classes. The most appropriate language for testing these groups of students will no doubt differ.

As discussed earlier in this chapter, SD/LEP questionnaire respondents were asked how each student would participate in NAEP if accommodations and adaptations were permitted. Based on responses to this question, two groups of students were identified: 1) those LEP students who would be tested in English, with or without accommodations; and 2) those LEP students who would be tested in their native language, with or without accommodations. Results on inclusion rates for these two groups of students are shown in table 5.8.

---

[4] National Center for Education Statistics. (1996). *Proceedings of the conference on inclusion guidelines and accommodations for limited English proficient students in the national assessment of educational progress.* (NCES Publication No. 96-861). Washington, DC: Author.

Table 5.8 — Percentages of limited English proficient students assessed, by prefered language of assessment, sample type and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

If accommodations/adaptations were available, how would this student participate in NAEP?	In English			In native language		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4						
S1[1]	50	88	12	56	56	44
S2[2]	109	78	22	87	8 ‡	92 ‡
S3[3]	116	90	10	68	66 ††	34 ††
Grade 8						
S1[1]	56	87	13	10	– –	– –
S2[2]	109	82	18	36	35	65
S3[3]	105	88	12	44	76 ††	24 ††
Grade 12						
S1[1]	21	91	9	1	– –	– –
S2[2]	136	81	19	16	– –	– –
S3[3]	111	91	9	9	– –	– –

[1] 1990-1996 operational inclusion critera

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

– – Indicates insufficient sample size to permit a reliable estimate.

‡ Indicates a significant difference between S1 and S2 results.

†† Indicates a significant difference between S2 and S3 results.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

The results suggest that procedural changes being made had their primary impact on inclusion rates at grades 4 and 8 among students who would be tested in their native language. First, these results suggest that, absent accommodations and adaptations, the revised criteria did result in lower inclusion rates, but that this occurred primarily at grade 4 among students who would be assessed in their native language. Grade-four inclusion rates in S2 were substantially lower than those in S1 (8 percent versus 56 percent) among students who would be assessed in their native language. S2 inclusion rates also appeared lower among grade-four LEP students who would be tested in English. However, this apparent difference is not statistically significant. Second, when the revised criteria were used, the provision of accommodations and adaptations increased inclusion at grades 4 and 8 primarily among students who would be tested in their native language. At both grades 4 and 8, inclusion rates were substantially higher in S3 (where accommodations were permitted) than in S2 (where they were not). Among LEP students who would be tested in English, the apparent differences were in the same direction (i.e., S3 had higher inclusion rates than S2), but were smaller and not statistically significant.

As shown in chapter 4, the majority of LEP students at all three grades are native speakers of Spanish. However, over 25 percent of LEP students at grades 4 and 8, and close to 50 percent at grade 12 are native speakers of some other language. Table 5.9 presents results on inclusion rates for two groups of LEP students: 1) those who are native speakers of Spanish and; 2) those who are native speakers of some other language. The results in table 5.9 suggest that the main impact of the revised criteria occurred for LEP students who are native speakers of Spanish. Among this group, inclusion rates were higher in S3 (where accommodations were provided) than in S2 (where they were not) at grades 4 and 8. At grade 12, the difference between S3 and S2 inclusion rates, though in the same direction, is not statistically significant. For LEP students with other native languages, the evidence is less convincing. At grades 4 and 12, S3 inclusion rates appear higher than those in S2, but the differences are not statistically significant. At grade 8, little difference is evident.

Table 5.9 — Percentages of limited English proficient students assessed, by native language, sample type and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

What is the student's first or native language?	Spanish			Another Language		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4						
S1[1]	107	64	36	15	– –	– –
S2[2]	178	35	65	59	67	33
S3[3]	159	73 ††	27 ††	52	88	12
Grade 8						
S1[1]	28	76	24	31	71	29
S2[2]	172	57	43	52	72	28
S3[3]	110	82 ††	18 ††	44	74	26
Grade 12						
S1[1]	16	– –	– –	14	– –	– –
S2[2]	108	75	25	48	72	28
S3[3]	46	84	16	81	87	13

[1] 1990-1996 operational inclusion critera

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

– – Indicates insufficient sample size to permit a reliable estimate.

†† Indicates a significant difference between S2 and S3 results.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

It appears that when the revised criteria were used accommodations and adaptations increased inclusion primarily for LEP students whose native language is Spanish.
Such differences are perhaps not surprising given the inclusion of the Spanish-bilingual version of the test. However, it is also surprising that when compared to the original inclusion criteria, the combined use of accommodations and adaptations with the revised inclusion criteria does not provide strong evidence of increased inclusion. At grades 4 and 8, the differences between S3 and S1 inclusion rates are in the expected direction, but are relatively small for both groups of students and are not statistically significant.

Inclusion Rates by Instructional Situation

As discussed in chapter 4, LEP students differ with respect to the type of subject matter instruction they receive. Some are mainstreamed and receive no special instruction. Others are receiving specially designed instruction in English (e.g., ESL classes, content-based ESL, or English-language instruction with some native language support). Still others are in bilingual education classes and are receiving substantial amounts of instruction in their native language. Table 5.10 presents results on inclusion rates for students by the type of mathematics instruction they were receiving in 1996.

Table 5.10 — Percentages of limited English proficient students assessed by type of mathematics instruction, sample type, and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

Type of mathematics instruction during the current year?	No special instruction			Specially designed instruction in English			Native language instruction		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4									
S1[1]	34	80	20	34	74	26	33	41	59
S2[2]	95	64	36	79	54	46	73	6	94
S3[3]	91	83 ††	17 ††	62	92 ††	8 ††	56	51 ††	49 ††
Grade 8									
S1[1]	32	86	14	28	57	43	6	– –	– –
S2[2]	129	80	20	47	27	73	24	23	77
S3[3]	77	82	18	73	75 ††	25 ††	14	—	—
Grade 12									
S1[1]	32	85	15	0	– –	– –	0	– –	– –
S2[2]	91	82	18	53	58	42	19	– –	– –
S3[3]	94	89	11	24	68	32	5	– –	– –

[1] 1990-1996 operational inclusion critera

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

– – Indicates insufficient sample size to permit a reliable estimate.

†† Indicates a significant difference between S2 and S3 results.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Under the revised criteria, the provision of accommodations and adaptations resulted in higher inclusion rates for all three groups of grade-four LEP students. For two of these groups, inclusion rates in S3 also appear higher than inclusion rates in S1. However, these latter differences are not statistically significant. S1 inclusion rates at grade 4 also appear higher than those in S2 for all three groups. However, these differences are also not statistically significant.

At grades 8 and 12, there is little to suggest that either of the procedural changes had an impact on inclusion rates for students receiving no special instruction. However, there is some evidence to suggest rates were increased for grade-eight students receiving specially designed instruction in English. At grade 8, inclusion rates appear higher in S3 than in either S2 or S1. However, only the S3/S2 difference is statistically significant.

As reported in chapter 4, almost all LEP students at grades 8 and 12 were receiving their mathematics instruction in English. At grade 4, the large majority also were receiving their instruction primarily in English. Table 5.11[5] presents results on inclusion rates for two groups of LEP students: 1) those receiving instruction at or above grade-level in English; 2) those receiving instruction below grade-level in English. The patterns of results differ somewhat by grade level.

At grade 4, the provision of accommodations and adaptations had an effect on inclusion rates for students receiving instruction below grade level. Among grade-four students receiving below-grade-level instruction, 78 percent of S3 LEP students were included compared to 28 percent of S2 LEP students. Inclusion rates in S3 also appear higher than those in S2 for students receiving instruction at or above grade level. This latter difference is not statistically significant.

[5] The reader should note that for native-language instruction the N's differ from those reported in Table 5.10. These differences are due to differing patterns of question-level omitting among questionnaire respondants.

Table 5.11 — Percentages of limited English proficient students assessed, by grade level of mathematics instruction in English, sample type, and grade: 1996 NAEP mathematics sample

	Grade level of mathematics instruction in English					
	At or above grade level			Below grade level		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4						
S1[1]	63	77	23	23	68	32
S2[2]	152	68	32	23	28	72
S3[3]	155	83	17	25	78 ††	22 ††
Grade 8						
S1[1]	45	91	9	21	42	58
S2[2]	157	66 ‡	34 ‡	28	57	43
S3[3]	122	88 ††	12 ††	35	58	42
Grade 12						
S1[1]	21	90	10	10	– –	– –
S2[2]	99	75	25	50	73	27
S3[3]	71	85	15	40	90	10

[1] 1990-1996 operational inclusion critera

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

– – Indicates insufficient sample size to permit a reliable estimate.

‡ Indicates a significant difference between S1 and S2 results.

†† Indicates a significant difference between S2 and S3 results.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

As has been the case for many of the analyses presented in this chapter, the combined impact of the revised criteria and the provision of accommodations does not result in clear increases in grade-four inclusion rates when compared to the use of the original criteria. Inclusion rates appear higher in S3 than in S1 for both groups of students. However, the magnitude of these apparent differences is smaller and is not statistically significant. Moreover, the data are unclear as to whether, absent accommodations and adaptations, the revised criteria are less inclusionary than the original criteria. Regardless of type of instruction, inclusion rates appear lower in S2 than in S1. However, neither of these apparent differences is statistically significant.

The results at grade 8 follow a different pattern. The impact on inclusion rates of providing accommodations and adaptations is only evident among students receiving grade-level instruction in English. Inclusion rates in S3 are higher than those in S2. Moreover, the inclusion rate for this group of students is also higher in S1 than in S2. This result suggests that, absent accommodations and adaptations, fewer of these students were included using the revised inclusion criteria than the original. Inclusion rates at grade 12 provide no strong evidence of the impact seen at grade 8. S3 inclusion rates appear higher than those in S2 for students receiving grade-level and below-grade-level English-language instruction. However, neither of these apparent differences is statistically significant.

Questionnaire respondents were asked to estimate the level of mathematics performance in English for their limited English proficient students. Inclusion rates by this estimated performance level are presented in table 5.12. Increases in inclusion rates were observed only in the grade-four groups. Inclusion rates were higher when accommodations and adaptations were provided. Once again, however, the combined use of the revised criteria and the provision of accommodations did not result in increased inclusion when compared to the original criteria. In the grade-eight and grade-twelve groups, patterns of inclusion rates were only slightly higher for the S3 than S2 groups, and the differences did not reach statistical significance.

Table 5.12 — Percentages of limited English proficient students assessed by estimated level of mathematics performance in English, by sample type and grade: 1996 NAEP mathematics sample

	Grade level of mathematics performance in English					
	At or above grade level			Below grade level		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4						
S1[1]	35	86	14	47	64	36
S2[2]	114	69	31	59	48	52
S3[3]	101	88 ††	12 ††	74	76 ††	24 ††
Grade 8						
S1[1]	19	– –	– –	44	69	31
S2[2]	71	79	21	67	57	43
S3[3]	55	91	9	50	59	41
Grade 12						
S1[1]	16	– –	– –	8	– –	– –
S2[2]	64	84	16	60	65	35
S3[3]	46	88	12	34	85	15

[1] 1990-1996 operational inclusion critera

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

– – Indicates insufficient sample size to permit a reliable estimate.

†† Indicates a significant difference between S2 and S3 results.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Chapter Summary

This chapter looked more closely at NAEP inclusion rates among LEP students from the mathematics assessment. In the first section of the chapter, results originally included in the *NAEP 1996 Mathematics Report Card for the Nation and States* were presented again and discussed in light of additional information available from the SD/LEP questionnaire.

In general, the percentages of LEP students assessed in NAEP without accommodation or adaptations exceeded questionnaire-based estimates of these same percentages. This suggests that increases in the numbers of LEP students included in NAEP are not likely to result solely from revisions to inclusion criteria that do not also involve the provision of accommodations and adaptations.

A comparison of the questionnaire respondents' estimates of the percentage of LEP students who should be legitimately excluded from participation in NAEP with the observed exclusion rates from NAEP indicates that further modest improvements in inclusion might still be possible if the list of permitted accommodations and adaptations can be expanded. Increasing the amount of permitted native-language testing (e.g., by providing alternate test forms in languages other than Spanish or allowing oral administration of the test in the student's native language) would be the most likely means of expanding inclusion.

In the second section of this chapter, inclusion rates among particular segments of the LEP population were discussed. The NAEP policy on including LEP students in the assessment based on the amount of English instruction received was changed for this assessment in two specific ways. First, the change involved a switch from using the number of years that the student attended a school where English was the primary language of instruction to using the number of years the student was receiving academic instruction in English. Second, the threshold for inclusion was changed from 2 years to 3 years. The second aspect of the change could be viewed as being somewhat less inclusionary.

Inclusion rates were analyzed by questionnaire-based reports of the length of time students were enrolled in English-language schools. These analyses provide some evidence that, when implemented without the provision of accommodations and adaptations, the revised criteria resulted in less inclusion among LEP students than did the original criteria. This evidence was strongest at grade 4. The analyses also showed, regardless of the inclusion criteria used, that the provision of accommodations and adaptations increased inclusion among grade-four LEP students enrolled in English-language schools for less than two years. Other results are less clear-cut. There is some evidence to suggest that the accommodations and adaptations may increase inclusion for all grades and students when the revised criteria are employed. However, providing accommodations and adaptations in conjunction with the revised criteria did not clearly result in higher inclusion rates than those achieved using the original criteria and allowing no accommodations.

Under the revised criteria, all students receiving academic instruction in English for three or more years were to be included in NAEP. Analyses based on questionnaire responses as to the number of years students were receiving academic instruction indicated that this ideal was not quite achieved. Inclusion rates among students with three or more years of academic instruction in English were high, but total inclusion was not achieved even where accommodations and adaptations were provided. Results also suggested that, under the new criteria, the provision of accommodations and adaptations did increase inclusion for all LEP students, regardless of the duration of their academic instruction in English. The evidence for this increase is most clear at grade 4, but similar patterns of results, though not statistically significant, are observed at grades 8 and 12. However, once again, the results suggest that the provision of accommodations and adaptations in combination with the revised criteria did not result in more inclusion than did the original criteria and no accommodations or adaptations.

Questionnaire respondents were asked to indicate how each student would participate in NAEP if accommodations and adaptations were permitted in English or in their native language. Results suggest that the procedural modifications being made to NAEP had their primary impact on inclusion rates at grades 4 and 8 among students who would be tested in their native language. Absent accommodations and adaptations, the revised criteria resulted in lower inclusion rates than did the original criteria at grade 4. When the revised criteria were used, the provision of accommodations and adaptations increased inclusion rates at grades 4 and 8 among students who would be tested in their native language.

In reviewing the impact on inclusion rates of testing in English versus Spanish, it appears that, when the revised criteria were used, accommodations and adaptations increased inclusion primarily for LEP students whose native language is Spanish. Such differences are perhaps not surprising given the inclusion of the Spanish-bilingual version of the test. However, it is surprising that, when compared to the original inclusion criteria, the combined use of accommodations and adaptations with the revised inclusion criteria provides little evidence of increased inclusion. At grades 4 and 8, the differences between S3 and S1 inclusion rates are relatively small for both groups of students and are not statistically significant.

When inclusion rates are examined by the type of subject matter instruction being given to students (no special instruction, specially designed English instruction, or bilingual instruction), the provision of accommodations and adaptations resulted in higher inclusion rates for all three groups of grade-four LEP students. In the two higher grades, the procedural changes had little impact on students receiving no special instruction. However, many of the differences are not statistically significant, resulting in a rather equivocal interpretation of the relationship among these variables.

LEP students were sorted into two groups based on questionnaire responses regarding grade-level of student instruction in mathematics: 1) at or above grade-level instruction in English, and 2) those receiving instruction below grade level in English. At grade 4, the provision of accommodations and adaptations had the most impact on inclusion rates among students receiving below grade-level instruction in English. At grade 8, the results followed a contrary pattern. The principal impact on inclusion rates occurred for students at/above grade level. As was the case for many of the analyses in this chapter, the combined impact of the revised criteria and the provision of accommodations provided no clear increase in inclusion over the original criteria.

Questionnaire respondents were asked to estimate student levels of mathematics performance in English. Two groups of students were formed based on these responses: 1) at or above grade-level performance; and 2) below grade-level performance. Using the revised criteria, the provision of accommodations and adaptations resulted in increased inclusion for both groups at grade 4. Once again, however, the combined impact of the revised criteria and the provision of accommodations provided no clear increase in inclusion over the original criteria.

Chapter 6

IRT Scaling and Model Fit of the Data

Introduction

Since 1984, NAEP has used Item Response Theory (IRT) methods to produce scale scores that summarize results in each assessed content area. One of the goals of this report is to determine whether the changes in testing conditions brought about by changes in inclusion criteria and presence of accommodations for special needs students in the 1996 assessments would result in scales that are equivalent to those used for the main NAEP reports. A possible negative implication for the NAEP program of increased inclusion of special needs students would arise if the performance data of these groups changed the 1996 scales in such a way as to invalidate comparisons with previous assessment years. For this reason, data from accommodated students were not aggregated with standard assessment data to produce the results reported in the *1996 NAEP Report Cards.*

Another goal is to determine whether by including more or different special needs students in NAEP, changes in the distributions of NAEP scale scores would result. Again, the production of such changes as a result of the methodological revisions could pose threats to the validity of trend comparisons of current NAEP results to those from previous assessments. This chapter describes analyses associated with the first goal. Chapter 7 addresses the second goal. All analyses in these two chapters were performed on data from the national assessment (i.e., no state assessment data were used in these analyses).

As an aid to understanding the analyses and results presented in this chapter, it may be helpful to distinguish two related, but somewhat different, technical questions. One question is related to assumptions of subgroup invariance that are inherent in IRT models. Prior to 1996, only students who could be meaningfully tested under standardized conditions were included in NAEP. As documented in earlier chapters, the provisions of accommodations and adaptations have, in some cases, increased the number of special needs students who are assessed in NAEP and, for the first time, resulted in assessment data collected under nonstandard conditions. The IRT methods used by NAEP assume that, to good approximation, a common IRT model fits the data for all the major subgroups reported on in NAEP. This assumption leads to an important psychometric question: Can the data obtained from special needs students tested with accommodations and adaptations be fit with the same IRT models as data obtained from students tested under standard NAEP conditions? The NAEP 1995 field test research and the work by Koretz[1] with state of Kentucky assessment data provide reason to question that assumption. However, neither study provided conclusive evidence and additional research on this issue was needed.

A related question is: What is the aggregate effect of such data on NAEP IRT scaling? In NAEP, a single IRT model is fit to the total data set available for a given assessment. This total data set has always included special needs students. Despite their increased numbers, data from special needs students represent a relatively small percentage of the total data set used in scaling. An even smaller percentage of the total data set is represented by special needs students who took the test with an accommodation or an adaptation. The presence of a small percentage of misfitting data may have little measurable impact on the meaning and technical characteristics of the aggregate NAEP scaling results. No previous research on these aggregate effects exists.

This chapter presents analyses pertinent to both of the above psychometric questions. The first question, whether common IRT models fit the data, was addressed through a series of Differential Item Functioning (DIF) analyses, and through studying plots of item fit IRT model/ data: The DIF and model fit analyses were carried out for items used in the accommodations, controlling for the overall level of proficiency. "Items used in the accommodations" refers to items appearing in special booklets that were prepared for students requiring certain types of accommodations. In this chapter, the special booklets are referred to as "accommodation booklets," and are described in more detail later. They contained the same items, in the same order, as one of the standard assessment booklets at each grade level in each subject. This allowed comparison of performance of the accommodated students with the standard assessment students on items taken under conditions that are as similar as possible, given the accommodations. In some parts of this chapter the items appearing in these booklets are referred to as the "accommodations items."

The second question concerning the aggregate effect of nonstandard procedures on NAEP scales was addressed by replicating the standard NAEP scaling procedures in the three different samples. That is, data from samples S1 (using 1992-94 student inclusion criteria and no accommodations), S2 (using revised criteria, but no accommodations), and S3 (using both the newer inclusion criteria and offering testing accommodations and adaptations) were independently scaled. Then comparisons were made of the technical characteristics of the resulting scales on a variety of item- and test-level measures of their technical characteristics.

This chapter is organized into three main sections. The first section provides a description of basic methodology as it applies to both the mathematics and the science assessments, and also contains the DIF analyses. The second and third sections present the findings from the separate IRT scalings in the two assessment subjects. Subsections contain descriptions of the specific methodology, comparisons of item fit statistics, and overall comparisons of how well the model fit the data in the different samples.

---

[1] Koretz, D. (1997). *The assessment of students with disabilities in Kentucky.* (CSE Technical Report No. 431). Los Angeles, CA: Institute on Education and Training, CRESST/RAND.

DIF Results for Accommodation Booklet Items for Science and Mathematics

DIF methods were used to compare item difficulties for standard assessment and accommodated students matched with respect to an overall measure of mathematics or science proficiency. As mentioned previously, these analyses are related to the issue of whether a common IRT model fits the data of both groups. DIF procedures, a standard NAEP analysis method for comparing performance of subgroups of students on each item of an assessment, were used.

Methods. Under the standard NAEP design, each student is randomly assigned one of several different booklets containing different combinations of blocks of items. In the 1996 NAEP design, at each grade level in both mathematics and science, one booklet was selected for administration to all accommodated students. Hence, information was gathered about performance of all accommodated students on the items in this booklet. If the typical NAEP design had been used, there would have been very small numbers of accommodated students responding to each item in the assessment. Using one accommodation booklet resulted in sufficient sample sizes to permit some comparisons of the performance of students requiring testing accommodations with that of students assessed under standard conditions. However, sample sizes were not sufficient to allow additional disaggregation. Thus, in the analyses discussed below, the group of students receiving accommodations includes students with disabilities as well as limited English proficient students. Moreover, the students in this group received varying kinds of accommodations and adaptations.

The DIF procedures compare item difficulty for two groups, typically referred to as the reference and the focal groups. Significance tests are provided to guide inferences about the differences emerging between the groups. Because all accommodated students took the same booklet, the DIF analyses for this report compared students assessed with standard procedures (who took that same booklet of items) to the accommodated group. In science, all students in S2 and S3 taking that booklet were selected, while in mathematics, all students taking that booklet in S1, S2, or S3 were selected for the DIF analyses.

The standard NAEP DIF analysis procedures lend themselves well to this type of comparison. In these analyses, mean performances of subgroups of students on each item in each booklet were compared, conditional on the overall booklet score. Taking into account these score means, accommodated students were compared with standard assessment students, while holding constant their ability levels in either mathematics or science. The Mantel-Haenzel procedure was used for dichotomously scored items while polytomously scored items (typically constructed-response items that receive partial credit scoring) were analyzed for DIF using the Mantel statistic. Both procedures are described in greater detail in the forthcoming *NAEP 1996 Technical Report*.[2]

[2] Allen, N.L., Carlson, J.E., & Zelenak, C.A. (forthcoming). *The NAEP 1996 technical report.* Washington, DC: National Center for Education Statistics.

The DIF procedures resulted in a sorting of items into one of three categories: (A) little or no indication of DIF; (B) weak indication of DIF; and (C) strong indication of DIF. Generally speaking, "category A" contains items with no statistically significant DIF. "Category B" items show statistically significant DIF, but relatively small effect sizes. For dichotomous items, the effect size is the difference in reference and focal group standardized item difficulties, as expressed in the ETS delta scale. For polytomous items, the effect size is the standardized mean difference divided by the item's standard deviation. Many such items encountered in field tests will eventually be included in operational test forms; category B items encountered in operational test forms are rarely dropped from the scoring of the test. "Category C" items exhibit statistically significant DIF and relatively large effect sizes. Such items encountered in field tests are typically not included in operational test forms. Category "C" items encountered in operational test forms are reviewed by a committee of trained test developers and subject matter specialists who are charged with making judgments as to whether or not the DIF is unfairly related to group membership. Items so judged are dropped from operational tests.

Science Accommodation Booklet DIF Results. The DIF results comparing accommodated administrations for the three grade levels in science are presented in tables 6.1 through 6.3, respectively. Each table shows the numbers of items classified into one of six categories, each of which is designated by a letter and a symbol. The letter portion of the designation indicates the DIF category (A, B, or C) to which the item was assigned. The symbol portion of the designation (+ or -), applicable only to B and C items, indicates the direction of the detected DIF. The "+" symbol indicates items that were differentially easier for the accommodations group, while the "-" symbol indicates items that were differentially difficult for the accommodations group. Counts are provided separately for dichotomous items[3], polytomous items[4], and all items combined. Grade-four per-item sample sizes were between 250 and 295 students for the standard administration group and between 140 and 148 students for the accommodations groups. Grade-eight per-item sample sizes were between 215 and 252 students for the standard administration group and between 120 and 149 students for the accommodations groups. Grade-twelve per-item sample sizes were typically between 151 and 199 students for the standard administration group and between 50 and 71 students for the accommodations groups.

As shown in table 6.1, ten of the 27 grade-four accommodations book items (37 percent) showed some indication of DIF, and all ten of these involved items that were differentially more difficult for the accommodations group. Only five of these ten items (18.5 percent of the items in the book) were "category C" items; however, four of these five were polytomous constructed-response items. As an aid to interpreting these results, it is helpful to compare these numbers and percentages to the operational DIF results for the 1996 assessment.[5] DIF analyses comparing White students and Black students, a set of DIF analyses similar to those presented here in that the focal group has somewhat small sample sizes, resulted in 10 percent of the total item pool being classified as "category B" or "category C" items (4 percent B items and 6 percent C items). Based on committee review, none of these DIF items was dropped. Thus, the amount of accommodation-related DIF observed for the grade-four science items appears to be greater than that encountered in the DIF analyses comparing Black students to White students.

---

[3] Items which may be scored right or wrong, for example multiple-choice and short, constructed-response items

[4] Items which require longer responses for which a student may give a correct, partially correct, or wrong answer

Table 6.1 — Distribution of items by DIF category for accommodations booklet for grade 4: 1996 NAEP science sample

	Dichotomous items	Polytomous items	Total
C+	0	0	0
B+	0	0	0
A	7	10	17
B–	4	1	5
C–	1	4	5
Total	12	15	27

NOTE: The symbol portion of the designation (+ or -), applicable only to B and C items, indicates the direction of the detected DIF. The "+" symbol indicates items that were differentially easier for the accommodations group, while the "-" symbol indicates items that were differentially difficult for the accommodations group.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Tables 6.2 and 6.3 present the DIF results for grades 8 and 12. At grade 8, only four of the 39 accommodations booklet items (about 10 percent) were categorized as B or C, and only one of these was a "category C item". White/Black DIF comparisons from operational NAEP analyses in the 1996 assessment also resulted in "category B" or "category C" classifications for ten percent of the grade-8 item pool (8 percent B and 2 percent C). Thus, at grade 8, there was little evidence of greater frequences of DIF related to the provision of accommodations than is typically encountered in operational DIF analyses with similar sample sizes. However, at grade 12, ten of the 36 items (28 percent) were categorized as B or C, though only two of these items (6 percent of the total) were "category C" items. White/Black DIF comparisons in the 1996 assessment resulted in "category B" or "category C" classifications for 13 percent of the grade-12 item pool (10 percent B and 3 percent C). At grade 12, similar to grade 4, the amount of accommodations-related DIF appeared greater than that encountered in White/Black DIF analyses in the 1996 assessment. In summary, accommodations - related DIF appeared greater than that encountered in White-Black DIF analyses (which had similar sample sizes) in grades 4 and 12, but not at grade 8.

[5] Allen, N.L., Carlson, J.E., & Zelenak, C.A. (forthcoming). *The NAEP 1996 technical report.* Washington, DC: National Center for Education Statistics.

Table 6.2 — Distribution of items by DIF category for accommodations booklet for grade 8: 1996 NAEP science sample

	Dichotomous items	Polytomous items	Total
C+	0	0	0
B+	1	0	1
A	14	21	35
B–	1	1	2
C–	0	1	1
Total	16	23	39

NOTE: The symbol portion of the designation (+ or -), applicable only to B and C items, indicates the direction of the detected DIF. The "+" symbol indicates items that were differentially easier for the accommodations group, while the "-" symbol indicates items that were differentially difficult for the accommodations group.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table 6.3 — Distribution of items by DIF category for accommodations booklet for grade 12: 1996 NAEP science sample

	Dichotomous items	Polytomous items	Total
C+	0	1	1
B+	2	0	2
A	10	16	26
B–	4	2	6
C–	0	1	1
Total	16	20	36

NOTE: The symbol portion of the designation (+ or -), applicable only to B and C items, indicates the direction of the detected DIF. The "+" symbol indicates items that were differentially easier for the accommodations group, while the "-" symbol indicates items that were differentially difficult for the accommodations group.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Mathematics Accommodation Booklet DIF Results. The DIF analysis results for the three grade levels in mathematics are presented in tables 6.4 through 6.6. Grade-four per-item sample sizes were between 250 and 295 students for the standard administration group and between 150 and 206 students for the accommodations groups. Grade-eight per-item sample sizes were between 283 and 333 students for the standard administration group and between 153 and 172 students for the accommodations groups. Grade-twelve per-item sample sizes were between 159 and 243 students for the standard administration group and between 59 and 72 students for the accommodations groups. At all three grades, there is little evidence of accommodations-related DIF.

As shown in table 6.4, six of the 27 grade-four accommodations book items (22 percent) showed some indication of DIF. Only two of these six items (7 percent of the items in the book) were "category C" items. NAEP 1996 mathematics assessment DIF analyses comparing White students and Black students resulted in 21 percent of the items subjected to DIF analyses[6] being classified as "category B" and zero percent classified as "category C." Thus, there is little evidence of accommodations-related DIF in mathematics at grade 4.

Table 6.5 presents the mathematics DIF results for grade 8. Seven of the 37 accommodations booklet items (about 19 percent) were categorized as B or C, and only one of these was a "category C" item. White/Black DIF comparisons from the 1996 mathematics assessment resulted in "category B" or "category C" classifications for 18 percent of the analyzed items (12 percent B and 6 percent C). Table 6.6 presents the results for grade 12. Six of the 43 items (14 percent) were categorized as B and none as C. White/Black DIF comparisons from the 1996 assessment resulted in B or C classifications for 15 percent of the analyzed grade 12 items (11 percent B and 4 percent C).

Table 6.4 — Distribution of items by DIF category for accommodations booklet for grade 4: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD NAEP

	Dichotomous items	Polytomous items	Total
C+	2	0	2
B+	2	0	2
A	19	4	23
B–	2	0	2
C–	0	0	0
Total	25	4	29

NOTE: The symbol portion of the designation (+ or -), applicable only to B and C items, indicates the direction of the detected DIF. The "+" symbol indicates items that were differentially easier for the accommodations group, while the "-" symbol indicates items that were differentially difficult for the accommodations group.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

[6] Only items newly introduced in 1996 mathematics were analyzed for DIF. All trend items had been previously analyzed during the 1990 and 1992 mathematics assessments.

Table 6.5 — Distribution of items by DIF category for accommodations booklet for grade 8: 1996 NAEP mathematics sample

	Dichotomous items	Polytomous items	Total
C+	1	0	1
B+	3	0	3
A	25	5	30
B–	3	0	3
C–	0	0	0
Total	32	5	37

NOTE: The symbol portion of the designation (+ or -), applicable only to B and C items, indicates the direction of the detected DIF. The "+" symbol indicates items that were differentially easier for the accommodations group, while the "-" symbol indicates items that were differentially difficult for the accommodations group.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table 6.6 — Distribution of items by DIF category for accommodations booklet for grade 12: 1996 NAEP mathematics sample

	Dichotomous items	Polytomous items	Total
C+	0	0	0
B+	3	0	3
A	30	7	37
B–	3	0	3
C–	0	0	0
Total	36	7	43

NOTE: The symbol portion of the designation (+ or -), applicable only to B and C items, indicates the direction of the detected DIF. The "+" symbol indicates items that were differentially easier for the accommodations group, while the "-" symbol indicates items that were differentially difficult for the accommodations group.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Summary of DIF Results

At two of the three grades in science there is evidence of accommodations-related DIF. Such evidence is noteworthy given the relatively small sample sizes involved in the accommodations analyses and, hence, the lack of power to detect statistical significance. These results are also consistent with those of the 1995 NAEP field test results,[7] as well as those reported by Koretz.[8] However, in contrast, the mathematics assessment provided little, if any, evidence of accommodations-related DIF. There are two noteworthy aspects of this pattern of results. First, sample sizes for the DIF analyses in mathematics were slightly larger than those in science. Other things being equal, one might have expected to observe more DIF in the mathematics results due to the presence of greater statistical power. That this expectation was not borne out suggests that other factors may be at work in producing the elevated levels of DIF in science.

One such factor may involve the availability of the bilingual booklet in mathematics at grades 4 and 8. A portion of the accommodations group in mathematics consists of native Spanish speakers who took the test with the bilingual book. All appropriately designed accommodations and adaptations, including the bilingual book, are intended to function as "corrective lenses," removing the distortion evident in scores obtained for students with disabilities and LEP students under standard testing conditions.[9] Perhaps the absence of this "corrective lens" in science is partly responsible for the higher observed levels of DIF. Of course, this interpretation is simply a conjecture. Myriad other factors may be responsible for the pattern of results. To give one example, there are differences in the numbers of items of different types in the mathematics and science assessments. The science assessment had more constructed-response items than the mathematics assessment. The science assessment also contained items based on hands-on science experiments and there were no comparable items in mathematics. Additional analyses are hampered by the heterogeneous nature of the accommodations group and the lack of ability to do meaningful analyses at lower levels of aggregation.

Science IRT Scaling Results

Item Response Theory (IRT) Procedures. Scales were produced for the three fields of science; 1) physical science; 2) earth science; and 3) life science. For each scaling analysis, item parameters were estimated using the standard NAEP BILOG/PARSCALE computer program, which is described in the forthcoming *NAEP 1996 Technical Report.*[10] A separate scaling analysis was conducted within each of the S2 and S3 samples. Additional scaling analyses were conducted within each of two random half samples of S2.

---

[7] Anderson, N.E., Jenkins, F.F., & Miller, K.E. (1996). *NAEP inclusion criteria and testing accommodations: Findings from the NAEP 1995 field test in mathematics.* Princeton, NJ: Educational Testing Service.

[8] Koretz, D. (1997). *The assessment of students with disabilities in Kentucky.* (CSE Technical Report No. 431). Los Angeles, CA: Institute on Education and Training, CRESST/RAND.

[9] McDonnell, L.M., McLaughlin, M.J., & Morison, P. (Eds.). (1997). *Educating one & all: Students with disabilities and standards based reform.* Washington, DC: National Academy of Science, National Research Council.

[10] Allen, N.L., Carlson, J.E., & Zelenak, C.A. (forthcoming). *The NAEP 1996 technical report.* Washington, DC: National Center for Education Statistics.

Random Sampling Differences in Science. By design, the S2 sample in science is approximately twice the size of each of the three mathematics samples and the S3 science sample. This feature of the 1996 NAEP design allowed for the computation of baseline comparison scaling results. By dividing the S2 science sample into two random halves and by replicating the comparative analyses in these two halves, estimates of the expected differences in scaling results between two equivalent samples of sizes similar to the other 1996 NAEP samples can be created.

Science Item Fit Statistic Distributions. A standard part of the results from the NAEP BILOG/PARSCALE computer program is an item fit statistic for each item. This statistic is a measure of how closely the responses of students conform to the theoretical item response function estimated for that item.[11] Lower numeric values of the fit statistic indicate better "fit." The fit statistic for any item cannot be used to determine unequivocally that an item does or does not fit the model because the exact form of the sampling distribution of these statistics is unknown. Consequently, under NAEP sampling and measurement design no tests of statistical significance may be conducted. It is reasonable to assume, however, that comparisons of the entire distributions of these statistics for the different samples might provide useful information on the relative fit of the IRT model to S2 data (which contained no accommodated sessions) and S3 data (where accommodations and adaptations were permitted).

For each scaling analysis, the fit statistic results for the items were categorized by size into five mutually exclusive categories. The resulting distributions of frequencies for the S2 and S3 samples in science are displayed in table 6.7, in which the categories are shown in a column labeled "Fit statistic category" for each grade level. The significance testing procedure is illustrated by using the grade-4 data from that table. The frequencies in the five categories for S2 can be used to estimate the proportions of fit statistics in those categories. A similar set of proportions can be estimated for S3. A chi-square test of homogeneity of proportions was used to test the hypothesis that the set of proportions across the categories for S2 is identical to the similar set for S3. The chi-square statistic (with 4 degrees of freedom) of 1.68 is not significant at the .05 level. Therefore, there is no strong evidence that the fit of the IRT model to the data is better in S2 than in S3.

Similar comparisons were made at the eighth and twelfth grades. The grade-12 comparison also resulted in nonsignificant differences between the distributions. The significance test for grade 8, however, indicates a difference. Examining the frequencies within the two columns, it appears that S3 had more items with fit statistics between 0 and .99 than S2, and the reverse was true for the highest category, 4 and larger. Contrary to expectations, however, these data suggest that the fit was actually worse in the sample comprised totally of standard assessment students (S2) than in the sample containing accommodated students (S3).

[11] Muraki, E., and Bock, R. D. *PARSCALE: IRT item analysis and test scoring for rating-scale data.* Chicago, IL: Scientific Software International, 1998.

Table 6.7 — Item fit statistic frequency distribution of items for S2 and S3: 1996 NAEP science sample

Fit statistic category	Sample		
	S2	S3	Total
Grade 4			
0.00 to 0.99	53	60	113
1.00 to 1.99	28	24	52
2.00 to 2.99	20	18	38
3.00 to 3.99	7	10	17
4.00 & larger	28	24	52
Total	136	136	272
	Chi-square = 1.68 not significant at the .05 level		
Grade 8			
0.00 to 0.99	41	58	99
1.00 to 1.99	37	43	80
2.00 to 2.99	36	29	65
3.00 to 3.99	15	22	37
4.00 & Larger	61	35	96
Total	190	187	377
	Chi-square = 12.49 significant at the .05 level		
Grade 12			
0.00 to 0.99	49	43	92
1.00 to 1.99	39	45	84
2.00 to 2.99	16	27	43
3.00 to 3.99	20	18	38
4.00 & Larger	62	53	115
Total	186	186	372
	Chi-square = 4.44 not significant at the .05 level		

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Judgments about the fit of the IRT Model to the Data

In addition to examining the distributions of fit statistics, the fit of the accommodation items to the model was examined graphically. This was done by plotting empirical and theoretical estimates of item response curves, for accommodated and standard assessment students, on the same graphs. The item response curve shows how the probability of getting each of the possible responses to an item changes as a function of students' scale scores. The theoretical curve, often referred to as the "item characteristic curve," (ICC) is computed from the IRT parameters estimated for the item. The empirical curve is determined from the posterior distribution of each student estimated at the conclusion of IRT parameter estimation by the PARSCALE/ BILOG computer program. The program uses a set of points on the proficiency scale, called "quadrature points," to estimate the item parameters. When convergence of estimation is reached, the proportion of students giving each response to the item at each of these points can be estimated from their posterior distributions. For example, for a dichotomous item, the proportions of correct and incorrect responses can be estimated. In addition, the proportions for each of two or more groups of students can be estimated. These data, plotted at the quadrature points, result in the empirical response curves.

Anderson, Jenkins, and Miller[12] describe procedures by which items were classified into three categories according to how well data fit a model. For this study, a similar procedure was used, although the categories are described slightly differently:

- **Nonlogistic** – the empirical item response function is judged not to be logistic in form and is not a good fit to the theoretical logistic IRT function estimated from standard assessment data;
- **Different parameters** – the empirical item response function is judged to be logistic in form, but with different parameters than those estimated from the standard assessment data; and
- **Good Fit** – the empirical item response function is judged to have a good fit to the theoretical logistic IRT function estimated from standard assessment data.

Items displaying severe nonlogistic fit are, theoretically, problematic. If the degree of misfit were judged serious, such items would be deleted from an assessment. The second category represents items that appear to be logistic in form, but have different parameters than the theoretical curve. This would result, for example, if an item were differentially difficult for the group from which the theoretical curve was estimated than for the group represented by the empirical curve.

Empirical ICCs for the accommodations booklet items were estimated for two groups of students — those who took the test with accommodations and those who took the test under standard conditions. These empirical ICCs were then plotted along with the operational theoretical ICCs for these same items (i.e., the curves estimated using the reporting sample data and used to report 1996 NAEP mathematics and science results). The plots were then sorted

[12] Anderson, N.E., Jenkins, F.F., & Miller, K.E. (1996). *NAEP inclusion criteria and testing accommodations: Findings from the NAEP 1995 field test in mathematics.* Princeton, NJ: Educational Testing Service.

into one of the three categories given above. The science results are presented in this section and the mathematics results are presented in a later section.

As noted earlier for item fit statistics, the sampling distributions of empirical ICCs are not well understood under NAEP's multistage student sampling and balanced-incomplete block item sampling designs. Thus, in comparing empirical ICCs for selected subgroups (e.g., students who took the assessment with accommodations) to reference curves estimated from the full data set, it is helpful to have some sense of the degree of fit one might expect to see for any randomly selected subgroup of similar size. Without such a backdrop, it is difficult to determine the degree to which any observed misfit is the result of between-group differences in ICCs or simply normal sampling variability. The inclusion of empirical ICCs for the group of students who took the same booklet as the accommodated students under standard conditions was intended to provide such a backdrop. However, it should be noted that this group is slightly larger than the group of accommodated students. Therefore, it is reasonable to expect closer concordance between empirical and reference plots for the standard conditions group.

As a point of clarification, it may be helpful here to sharpen the distinction between the reporting sample data that generated the operational ICCs (the reference ICCs given in the plots) and the standard administration data used to generate the empirical ICCs. The data from which the operational ICCs were estimated came from students in the S2 sample who were administered the item. Since accommodations and adaptations were not permitted in S2, all of these students took the test under standard conditions. Furthermore, since the blocks of items in the accommodations booklet also appeared in other NAEP booklets, these data include students who have taken booklets other than the accommodations booklets. The data from which the empirical ICCs were generated came from the students in both the S2 and S3 samples who took the accommodations booklet under standard conditions. Thus, a part of the data (the part from S2) was used to estimate the operational item parameters, while the remainder of these data (the part from S3) was not.

In the case of the present analysis, empirical ICCs were obtained for two groups of students: students who took the booklet with accommodations or adaptations, and students who took the booklets under standard conditions. As with the DIF analyses presented above, the former group consists entirely of students with disabilities and LEP students from the S3 sample. The latter group consists of all students in the S2 and S3 samples, including students with disabilities and LEP students, who took the test under standard conditions. Per-item sample sizes for these analyses are identical to those for the DIF analyses discussed in this chapter.

As noted earlier, each of these plots was sorted into one of the three categories given above. Science results are provided in table 6.8. The percentages of items sorted into each of the three categories are shown separately for accommodated and standard assessment students. As with analyses in the previous section, a chi-square test of homogeneity of proportions was used to test the hypothesis that the percentage of items falling in each of the three categories was identical for the accommodated and standard assessment students.

As can be seen in table 6.8, the chi square statistic (with 2 degrees of freedom) was significant at the .05 level for the fourth and twelfth grades but not at the eighth-grade level. Examining the grade-4 frequencies, it appears that most of the differences occur between the "Different parameters" and "Fit with standard parameters" columns. This suggests that,

Table 6.8 — Science accommodation booklet percentages of IRT items by fit categories

Sample	Judged type of IRT curve fit		
	Nonlogistic	Different parameters	Fit with standard parameters
Grade 4			
Accommodated students	19	54	27
Standard assessment students	16	19	65
	Chi-square = 32.73 significant at the .05 level		
Grade 8			
Accommodated students	27	68	5
Standard assessment students	32	57	11
	Chi-square = 3.64 not significant at the .05 level		
Grade 12			
Accommodated students	34	54	11
Standard assessment students	20	37	43
	Chi-square = 25.77 significant at the .05 level		

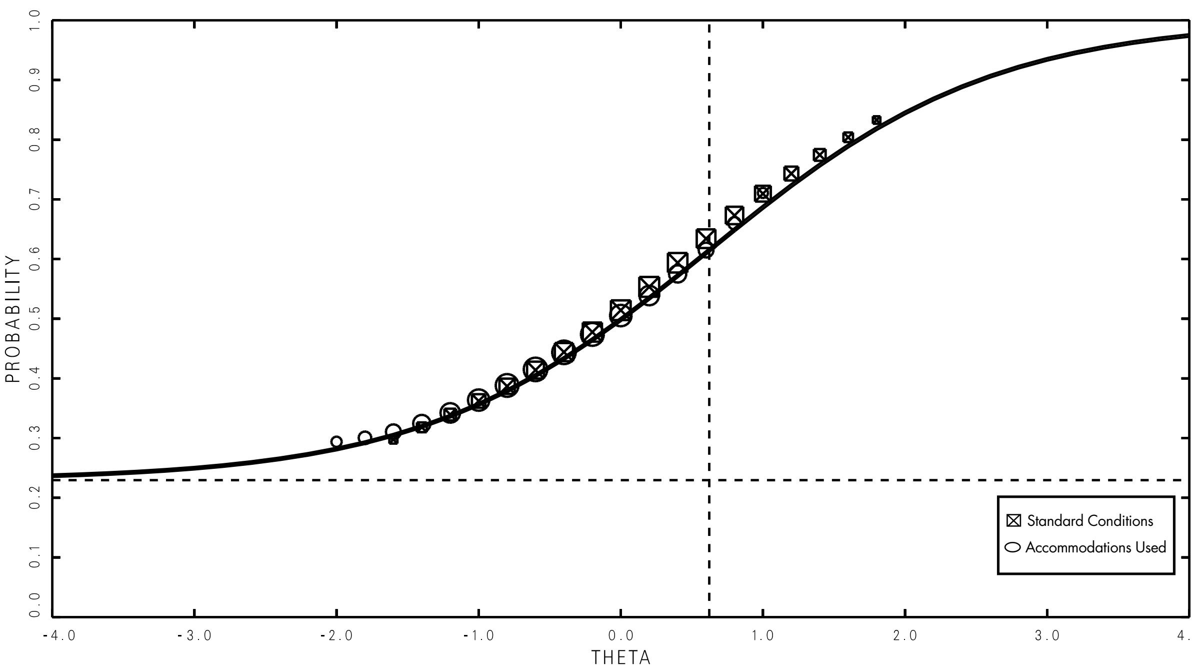
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

although similar IRT models might fit the data from both samples, the item parameters may differ. At grade 12, differences are evident in all three categories. Because of the extremely small sample sizes, especially compared to those used to fit the IRT models in operational NAEP, it is difficult to reach conclusions from these analyses. However, the results of the judgments of item level IRT model fit are consistent with the earlier presented DIF analyses. They suggest that, for at least some items at grades 4 and 12, a common ICC cannot be fit to the data for accommodated and standard assessment sessions.

Figure 6.1 provides some examples of item response curves for the 1996 science and mathematics data that were judged to be in each classification. The *x*-axis in these plots, labeled "THETA," represents the field of science or mathematics content strand scale in standard units (mean zero, standard deviation one). Plots a) and b) show curves for dichotomous items for which the data from students assessed under standard conditions (squares with X in them) fit the standard assessment item parameters well. In plot a) the data from the accommodated students (circles) fit well, but in plot b) their empirical ICC would be better fit with a logistic function with different parameters (about the same slope, but a higher difficulty parameter). Plot c) represents a polytomous item for which it appears that a logistic IRT model would fit the empirical data with different parameters than the standard assessment for both groups. Plots d) and e) demonstrate a dichotomous and a polytomous item for which the empirical data do not fit logistic curves for either accommodated or standard assessment students.

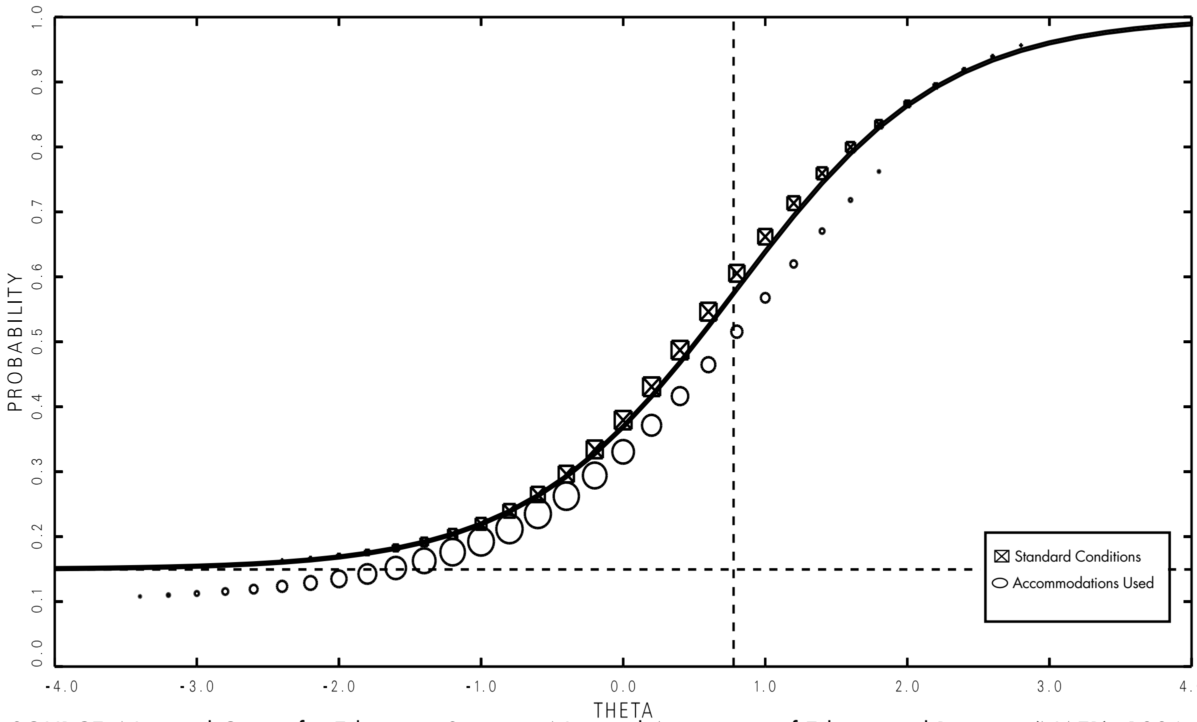
Figure 6.1 — Five examples of item response functions for accommodated versus non-accommodated students

a) Item response function for a dichotomous item in which standard assessment students fit well



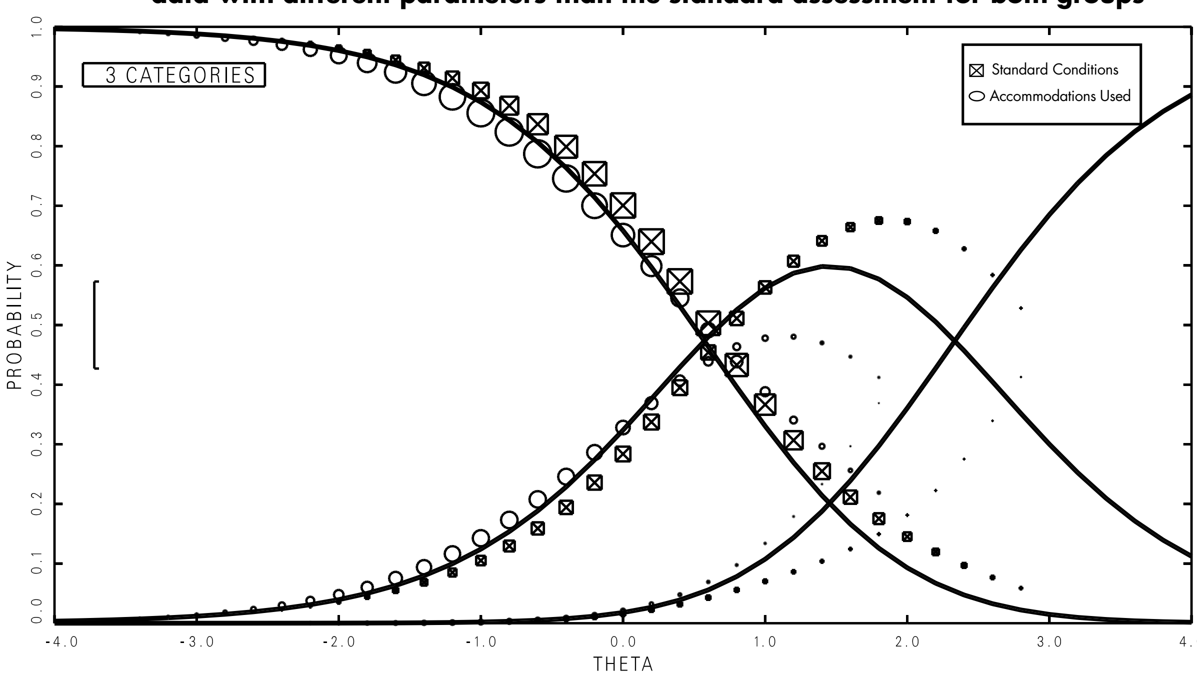
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

b) Item response function for a dichotomous item where accommodated students would fit a logistic function with different parameters



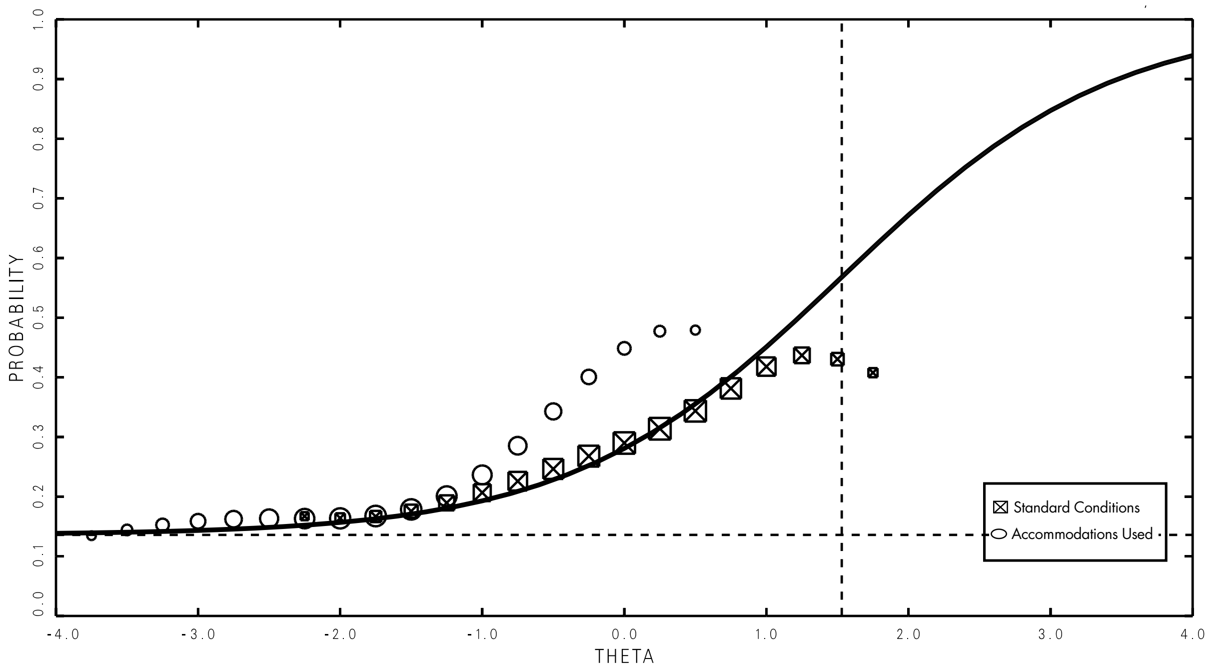
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

c) Item response function for a polytomous item where IRT model would fit the data with different parameters than the standard assessment for both groups



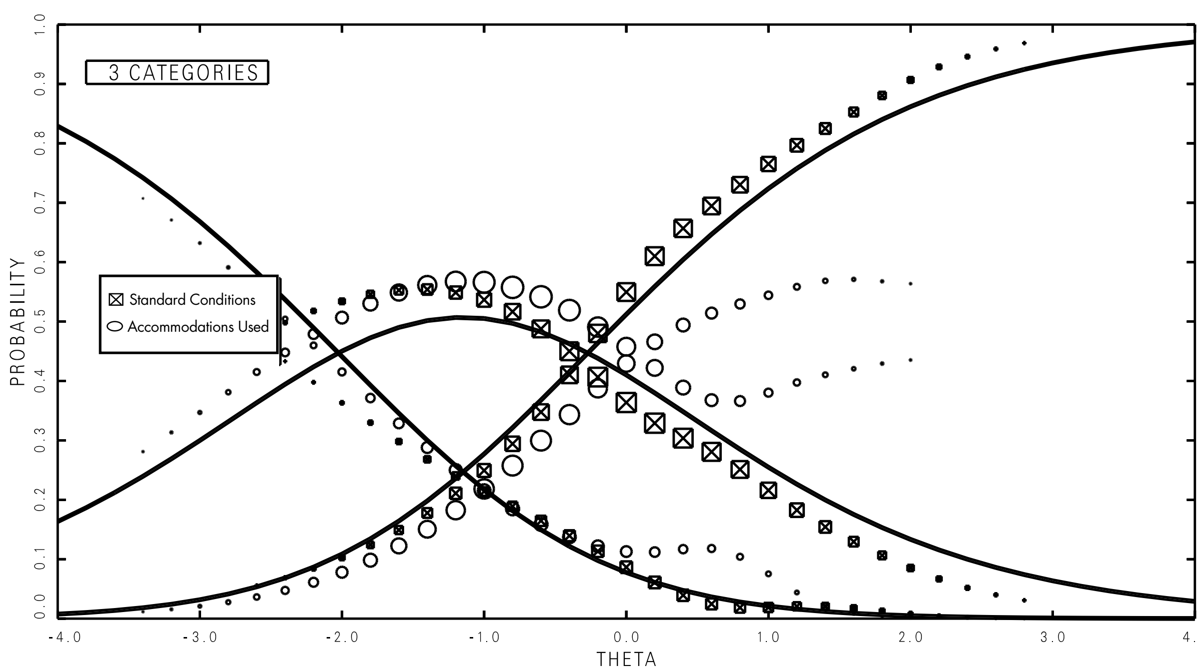
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

d) Item response function for a dichotomous item for which empirical data do not fit logistic curves for either accommodated or standard assessment students



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

e) Item response function for a polytomous item for which empirical data do not fit logistic curves for either accommodated or standard assessment students



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Overall Comparisons of the Item Response Theory (IRT) Models for Accommodated Students and Students Assessed Under Standard Conditions

The item response curves shown on the previous pages provide examples of the fit of the IRT model to the item response data collected from two groups of students (accommodated and standard assessment) on one set of items. Whether the actual scales, which are based on all items, are affected by changes in inclusion rules and presence of accommodations is a different question. The large number of item parameter estimates to be compared for each analysis made it necessary to condense the results of the IRT analyses to a summary form. Overall comparisons are made by computing the test characteristic functions and test information functions[13] for each of the scales at each grade level. As mentioned above, these functions are based on estimates of the parameters of all items on the scale. They describe some of the basic psychometric characteristics of the assessment instrument. The test characteristic function expresses the relationship between the proportion of items an examinee is expected to answer correctly and the position of the student on the proficiency scale for that field of science or mathematics content strand scale. Hence, it is a conditional measure: a measure of the expected proportion correct, conditional on the proficiency value. The test information function is also conditional. It expresses the precision of a maximum likelihood estimate of a student's proficiency on the scale, conditional on the actual proficiency value. Although NAEP does not use such maximum likelihood estimates, it was felt that the results would provide a useful estimation of the precision of measurement across possible proficiency values on each scale.

In general, an IRT scaling of sample data results in an arbitrary metric. When two or more samples are scaled separately, each sample scale is in a different arbitrary metric. In order to make comparisons between samples (S1, S2, and S3) it was necessary to transform the scales to a common metric. This was done within scale and within grade. That is, within each scale for each grade, the results for each sample were transformed to a common metric using the generalized Stocking-Lord procedure used in some past NAEP analyses. The methodology is documented in the *NAEP 1996 Technical Report*.[14]

Test characteristic curve results are reported in this chapter in subsequent sections for science and mathematics. Similarly, test information curve results are also reported for science and mathematics. Additional results are provided in appendices C through F.

It is important to note that these analyses, unlike those shown on the previous pages, include all items in the assessment scale, not just the items in the accommodation booklet. The reason for this is that the scales on which NAEP results are reported are developed including all items in the assessment. In theory, the effects of changing the conditions of a single item in an assessment (sampling, administration procedures, wording of an item) could change the scale. With a large number of items, however, the effect of changes to a single item would usually be

---

[13] Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[14] Allen, N.L., Carlson, J.E., & Zelenak, C.A. (forthcoming). *The NAEP 1996 technical report.* Washington, DC: National Center for Education Statistics

negligible. This section investigates the question of whether changes on some items for some students (accommodation items) had any effect on the overall scaling results. Hence, it was necessary to include all of the items in this analysis.

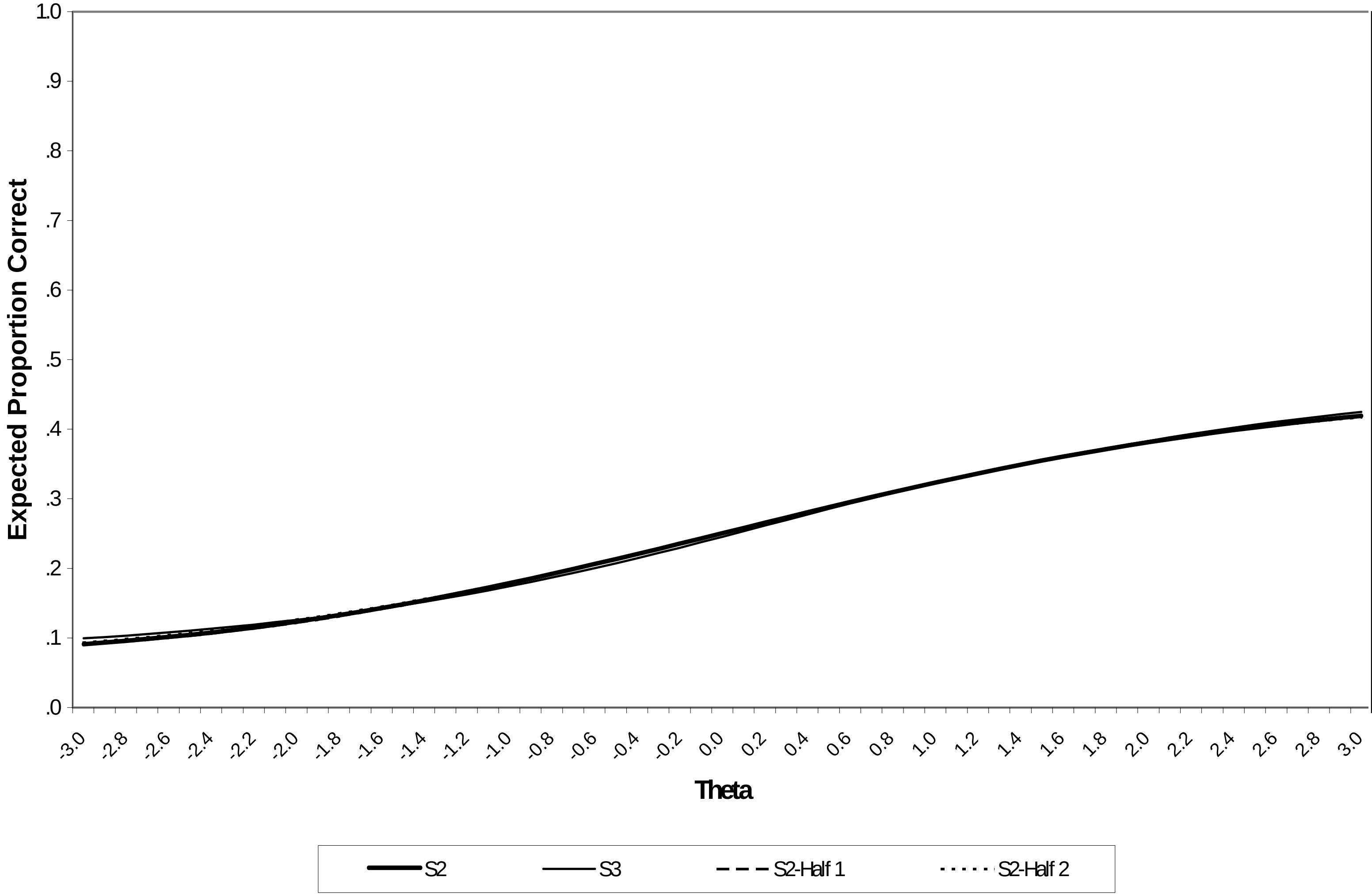
Science Scaling Procedures. As mentioned previously, the science assessment comprises three scales at each grade level, based on the three fields of science: physical science, earth science, and life science. Because many of the results are highly similar, examples, rather than complete results for each scale, are presented in each section. Additional results are presented in appendices C and D.

Scaling models were fit to the data separately for each of the three science scales, for each of the three grades, and for each of the two sample types (S2, 1996 inclusion rules with no accommodations; S3, 1996 inclusion rules with accommodations). In addition, as mentioned earlier, because the S2 sample is about twice the size of the S3 sample (and also about twice that of each of the three mathematics samples), an additional type of comparison of model fit was made — between two random halves of S2. The result is a total of 36 different scaling analyses. Because the S2 sample was the national reporting sample, however, item parameters already existed for that sample, so this report required no new scaling of this sample at each grade level.

Science Test Characteristic Curves. In this section the similarity of the measurement scales in the two science samples, and the two random-half samples, is illustrated by comparing the test characteristic curves at each grade level.

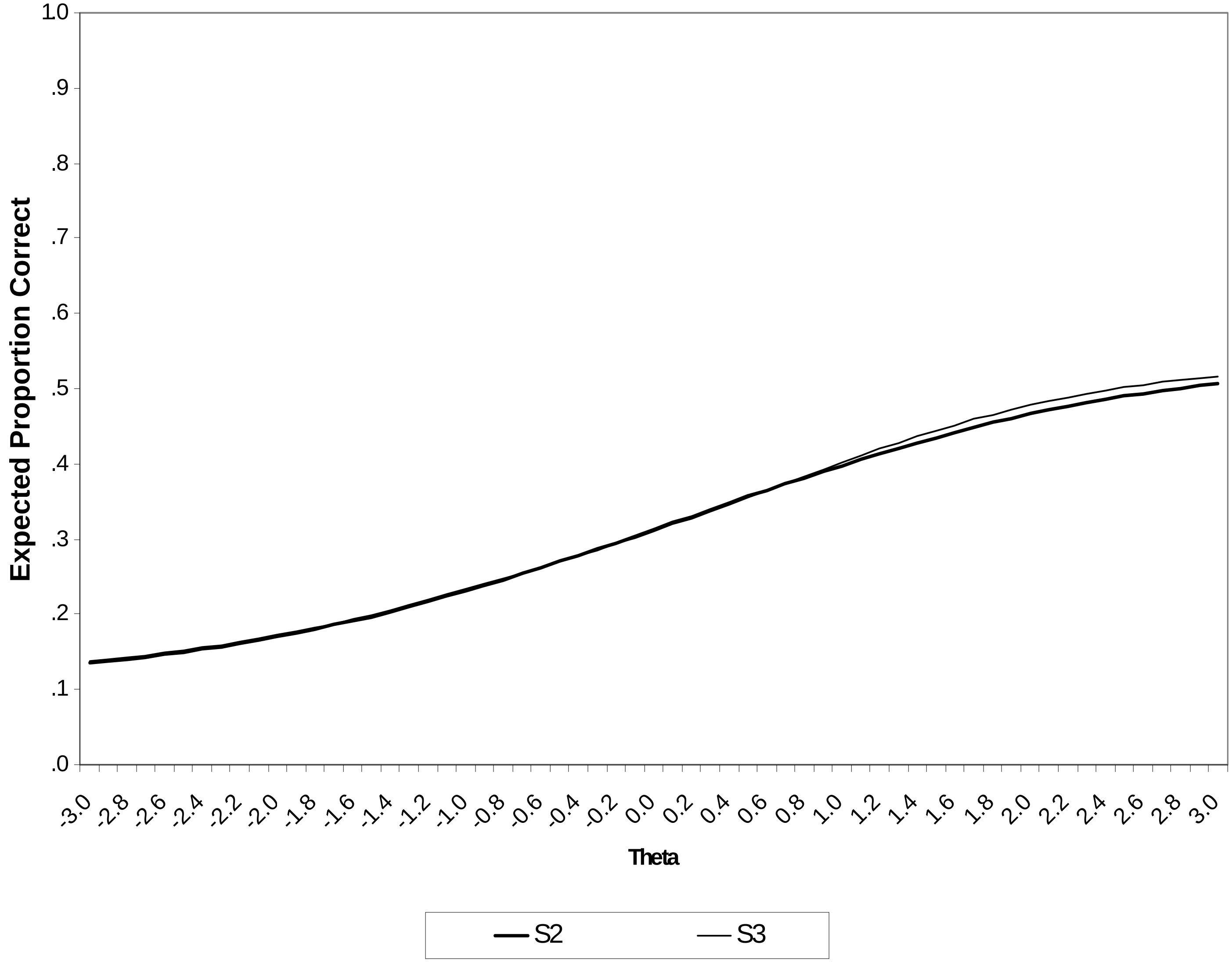
The test characteristic function indicates the expected level of test performance for students with varying degrees of proficiency. It also indicates how well the measurement instrument differentiates between students having slightly different values on the scale. Test discrimination is sharpest along those sections of the scale where the test characteristic curve has the steepest slope. This type of differentiation is usually referred to as discrimination of the test. Examples of test characteristic curves for the fourth-grade physical science scale are presented in figure 6.2. Figure 6.2 contains plots of the test characteristic curves for the full S2 sample, the two S2 half-samples, and the S3 sample. The proficiency scale in the plot was taken from the PARSCALE/BILOG result (mean near zero and standard deviation near one) after the results for the samples were transformed to a common metric, but before transformation to the final 1996 NAEP science reporting metric (mean 150 and standard deviation 35). The curves for the S2 and S3 samples, and for the two S2 half samples, are nearly indistinguishable. Figure 6.2 shows that the fourth-grade physical science scale has nearly identical expected performance levels and discriminates almost exactly the same for IRT scales fit to the student response data in the two different samples as well as in the two random half samples.

Figure 6.2 — Test characteristic curves for grade-4 physical science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

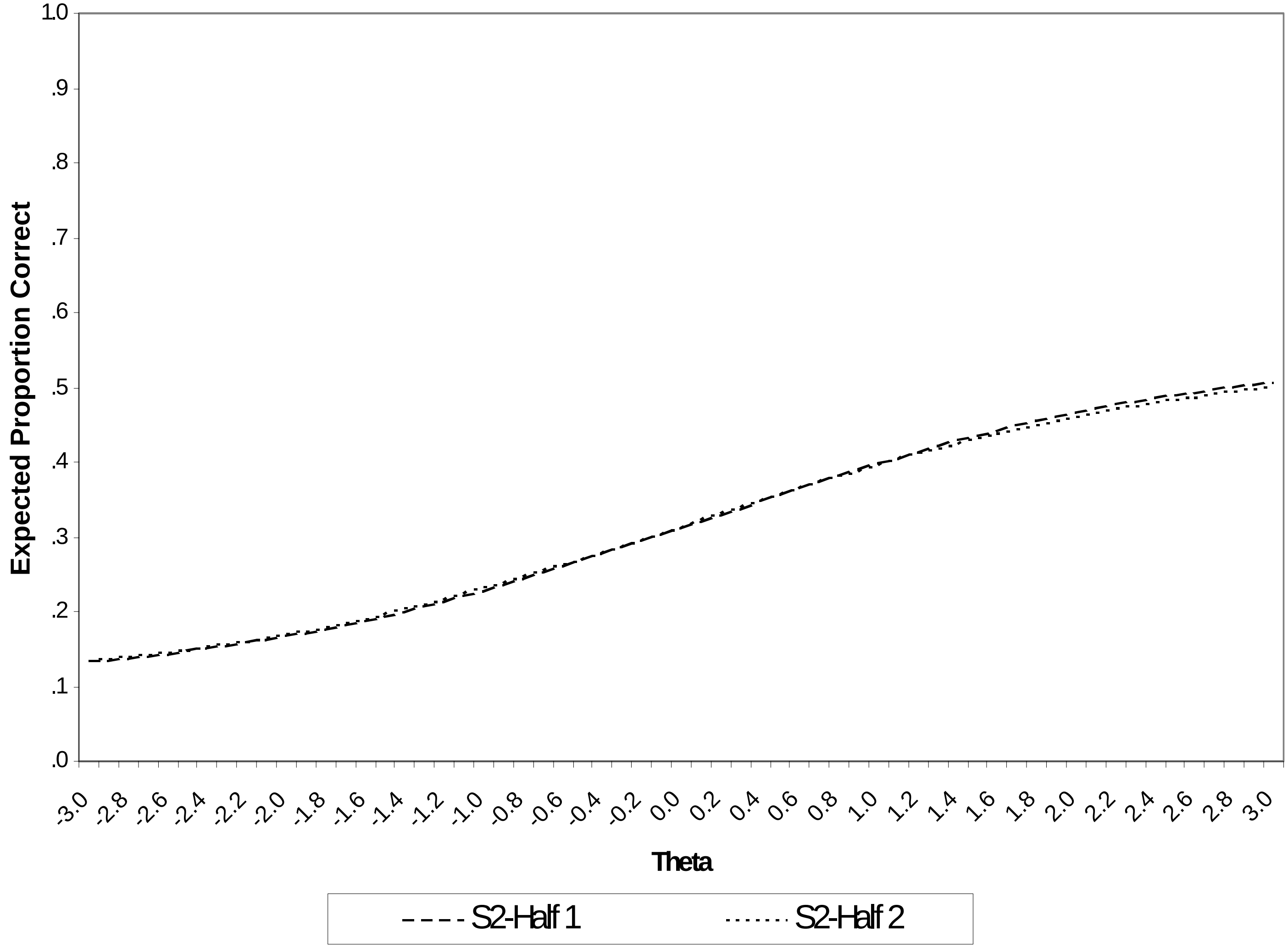
Figure 6.3 — Test characteristic curves for grade-4 earth science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

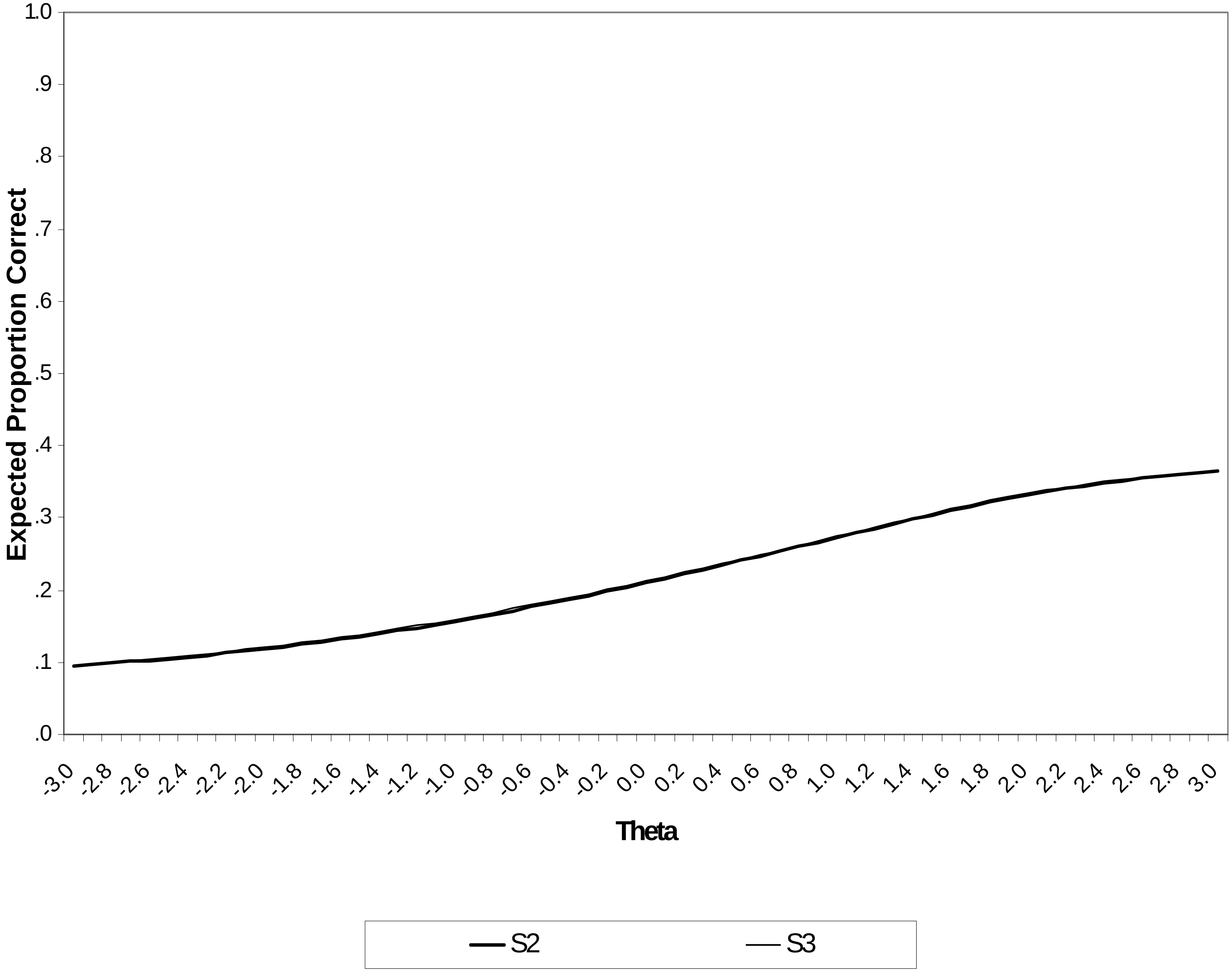
Figure 6.4 — Test characteristic curves for grade-4 earth science scale – S2 random halves

THE NATION'S REPORT CARD naep



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

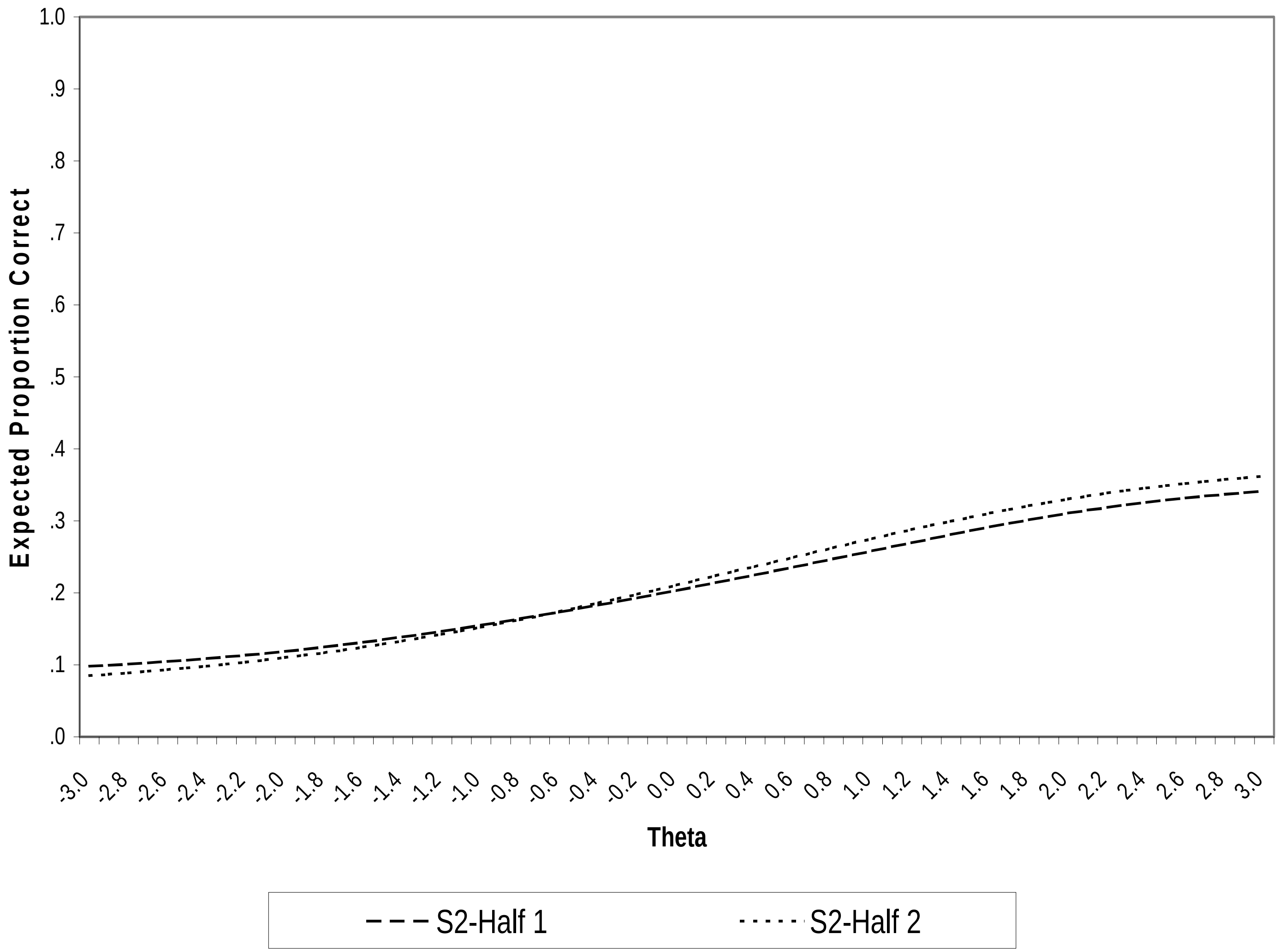
Figure 6.5 — Test characteristic curves for grade-4 life science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figure 6.6 — Test characteristic curves for grade-4 life science scale – S2 random halves

THE NATION'S REPORT CARD



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figures 6.3 through 6.6 display test characteristic curves for the other fourth-grade science scales. Those for the eighth and twelfth grades are presented in appendix C. In order to make comparisons of interest more readily, only two curves are shown in each of these figures, rather than four (as in figure 6.2).

Comparison of figures 6.3 and 6.4 (grade-4 earth science) with figures 6.5 and 6.6 (grade-4 life science) provides the basis for an overall summary of the results for the different types of samples. In one instance, the differences between the curves of S2 and S3 in earth science (figure 6.3) is larger than that between the two random-half samples (figure 6.4), particularly for theta greater than one. However, in another instance, the curves for S2 and S3 in life science (figure 6.5) are practically identical, whereas those for the two random-half samples (figure 6.6) actually differ more than those in figures 6.3 or 6.5. Similar results may be seen in the grade 8 and 12 curves in figures C1 through C7 of appendix C. On balance, the random sampling differences in the curves for the two random halves of S2 were as great as the differences between the S2 and S3 curves and it seems reasonable to conclude that the IRT scales are similar for the S2 and S3 samples.

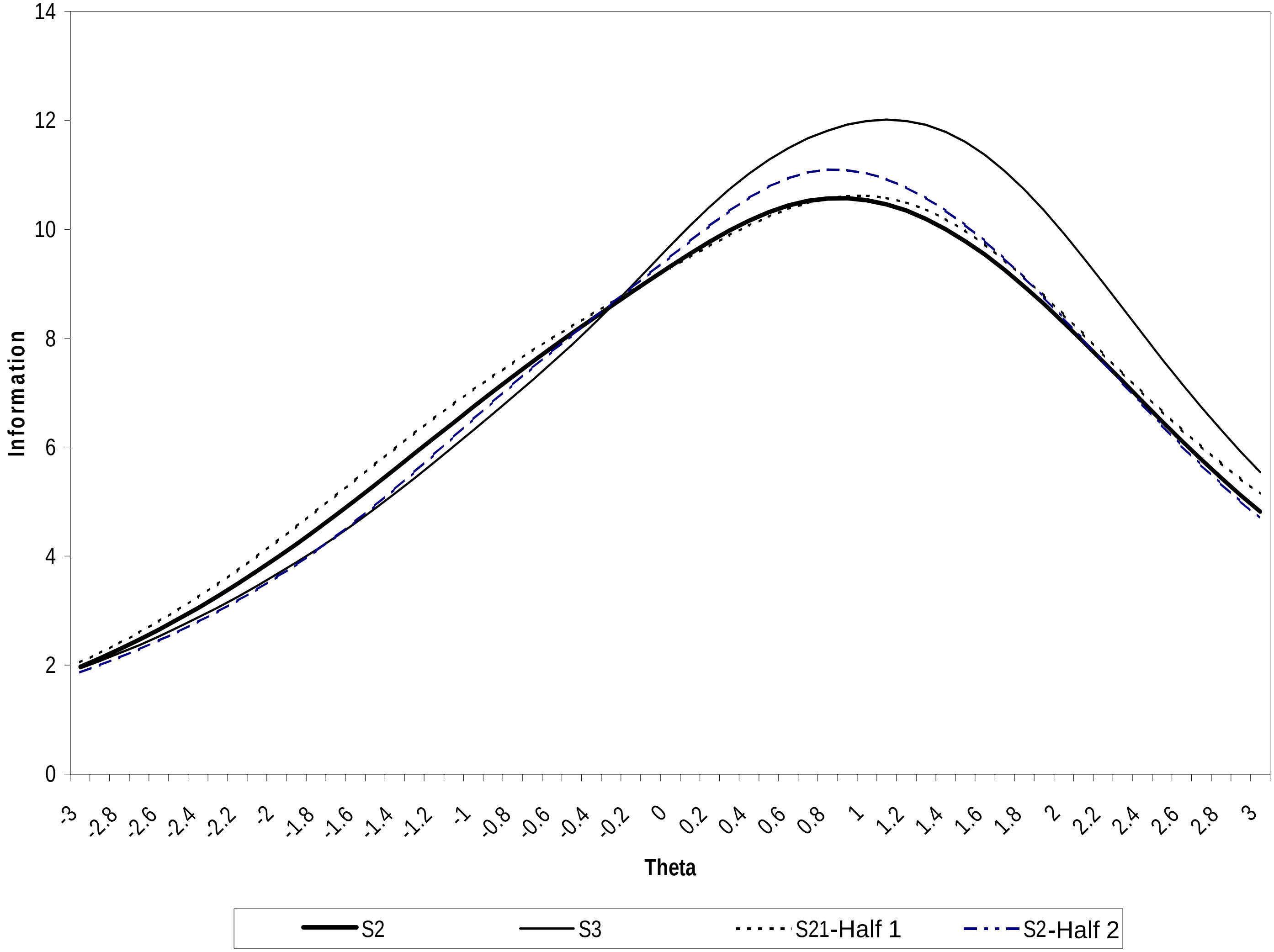
Overall, the differences between test characteristic curves based on independent estimates of item parameters within S2 and S3 are extremely small. Hence, there is no evidence that the presence of data from the accommodated students has changed the nature of any of the scales at any grade level.

Science Test Information Curves. The test information function expresses the relationship between precision of measurement and points on the scale. More information represents greater precision. Test information provides data about a test similar to that provided by the standard error of measurement. Although many test publishers report only a marginal (overall) standard error statistic, separate values of the standard error can be estimated conditional on scores on the test. In other words, it is always possible to estimate standard errors for different score points (or ranges of score points) on a measurement scale. Such information is more useful than one overall statistic because it tells the user how precise the estimates of proficiency are at different points on the scale. Information is proportional to the reciprocal of the square of the conditional standard error. Hence, a larger value of information indicates measurement that is more precise. For instruments scaled using IRT models, information is generally reported in a curve showing the differential precision of measurement over the range of proficiency scores.

Examples of test information function curves for the fourth grade are presented in figures 6.7 through 6.9. Those for grade-8 and grade-12 science are presented in appendix D. Although the differences between the different curves appear to be larger than those in the previously discussed test characteristic curves, this is at least partially a function of metric. Information is in a variance-like metric and hence it is related to squared deviations.

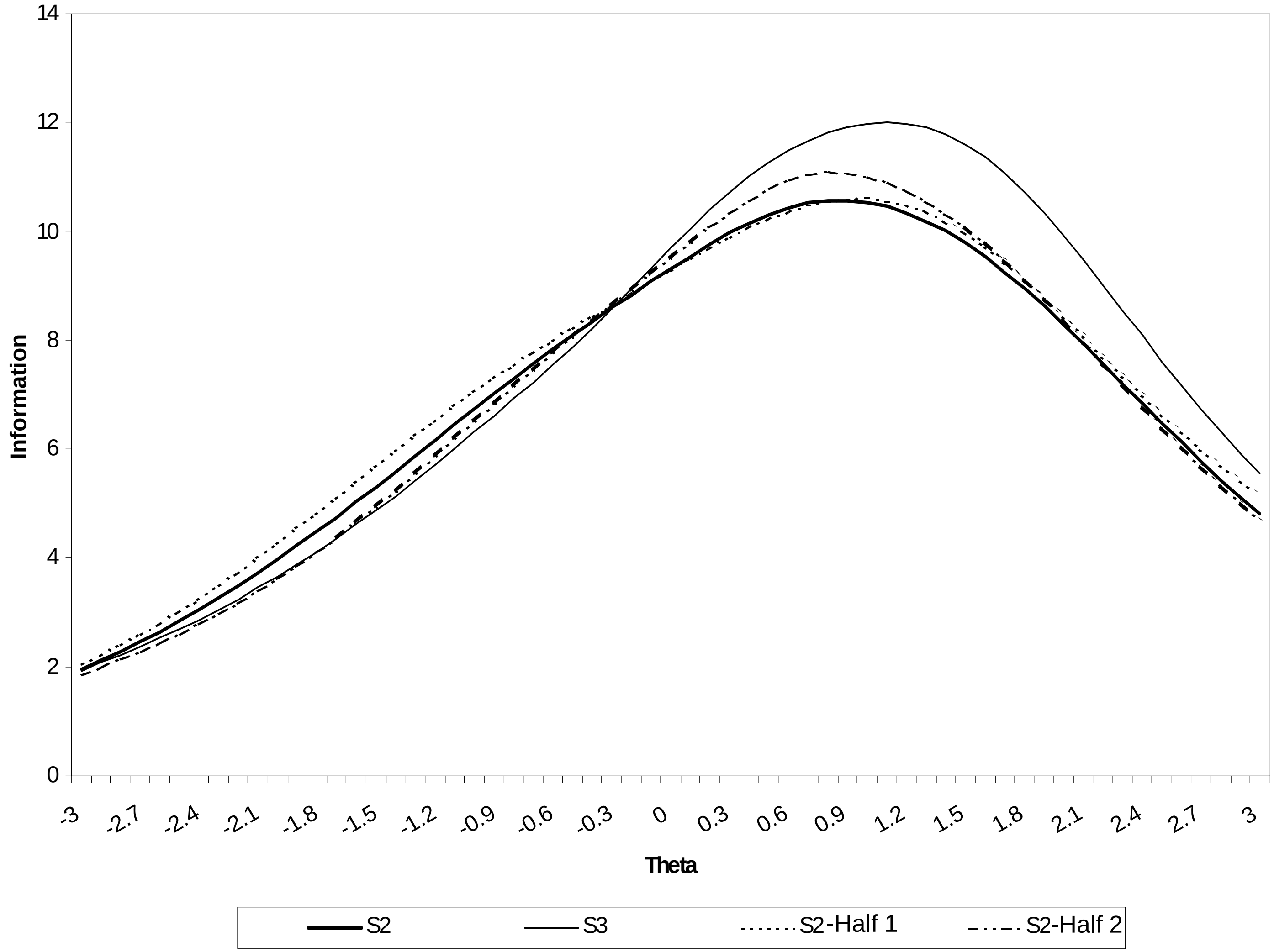
THE NATION'S REPORT CARD naep

Figure 6.7 — Test information curves for grade-4 physical science scale



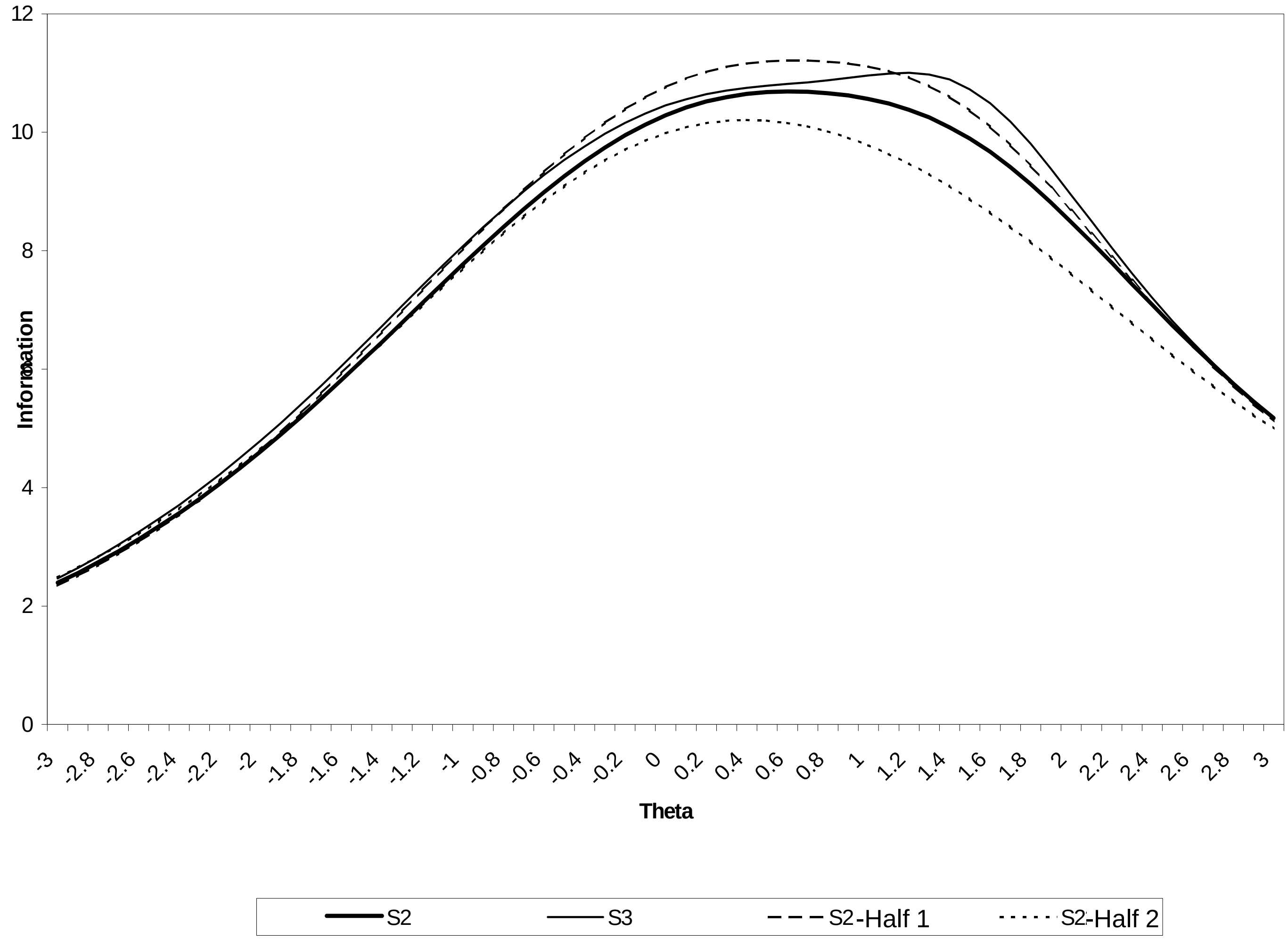
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figure 6.8 — Test information curves for grade-4 earth science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figure 6.9 — Test information curves for grade-4 life science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Squared values of a variable (for values greater than one), of course, are expected to result in the emphasizing of differences when compared to those based on a non-squared metric.

Another important factor in interpretation of information functions is the number of items making up a scale. In general, one expects to observe greater sample-to-sample variability in information functions for scales made up of few items.

Some of the figures (figures 6.7 and D.6) show larger differences between the S2 and S3 curves than between the curves for the two half samples. For example, for the grade 4 physical sciences scale (figure 6.7) the differences in the S2 and S3 information functions appear large, relative to the size of the differences between the half-samples of S2, above theta values of 0. On the other hand, other scales (figures 6.8 and D.1 through D.5) show as much, or more (figure 6.9), difference between the two random half samples as between the S2 and S3 samples. For example, for the grade 4 life science scale (figure 6.9) the differences between the information functions for the half-samples appear large, relative to the size of the differences between the S2 and S3 sample above theta values of 0.

Overall, the information curves show little evidence of major differences in precision of measurement depending on whether or not accommodations are offered to the students in the science assessment.

Mathematics IRT Scaling Results

The mathematics assessment comprises five scales at each grade level, based on the five content strands: number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and, algebra and functions. Because many of the results are highly similar, examples, rather than complete results for each scale, are presented in each section. Additional results are presented in appendices E and F.

Mathematics Scaling Procedures. Separate analyses were conducted within each of the three sample types (S1, original inclusion criteria with no accommodations; S2, revised inclusion criteria with no accommodations; and S3, revised inclusion criteria with accommodations) at each grade level.

The reader should note one important difference between the science results and the mathematics results in this section. In the 1996 NAEP science assessment, the reporting sample was identical to S2, so that sample was discussed in the section on science IRT scaling results. As previously mentioned, however, the mathematics reporting sample for 1996 was a combination of parts of S1 and S2. Separate results for these two mathematics samples are discussed in most of this section.

Table 6.9 — Item fit statistic frequency distribution of items for S1, S2, and S3: 1996 NAEP mathematics sample

Fit statistic category	Sample			Total
	S1	S2	S3	
Grade 4				
0.00 to 0.99	99	96	90	385
1.00 to 1.99	21	19	23	63
2.00 to 2.99	9	10	17	36
3.00 to 3.99	6	10	4	20
4.00 & larger	9	9	10	28
Total	136	144	144	432
	Chi-square = 6.86 not significant at the .05 level			
Grade 8				
0.00 to 0.99	110	100	107	317
1.00 to 1.99	31	34	33	98
2.00 to 2.99	11	13	9	33
3.00 to 3.99	5	9	3	17
4.00 & larger	7	8	12	27
Total	164	164	164	492
	Chi-square = 6.22 not significant at the .05 level			
Grade 12				
0.00 to 0.99	114	113	96	323
1.00 to 1.99	18	24	45	77
2.00 to 2.99	12	9	14	35
3.00 to 3.99	7	9	6	22
4.00 & larger	14	10	14	38
Total	165	165	165	495
	Chi-square = 10.26 not significant at the .05 level			

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

The IRT scaling models were fit to the data separately for each of the five scales, each of the three grades, and each of the three sample types. The result is a total of 45 different scaling analyses.

Mathematics Item Fit Statistic Distributions. Distributions of the item fit statistics were compiled for the three sample types at each grade level, and are presented in table 6.9.

The analyses performed on these data were similar to those for science reported in table 6.7. The table is divided into sections by grade. For each grade, the frequencies of items having fit statistics in the five ranges indicated in the first column are provided by sample type (S1, S2, and S3). Chi-square tests were conducted to test the hypothesis that the proportions underlying these frequencies are the same in the populations represented by the three samples.

At all three grade levels, the statistical tests indicated no significance. It would appear that the two different sets of inclusion criteria, and the presence or absence of accommodations, did not change the fit of the mathematics items to the model, at least in the sense of fit represented in the distributions of the fit statistics.

Judgments About the Fit of the IRT Model to the Data. As was done for science, empirical ICCs for the accommodations booklet items were estimated for two groups of students — those who took the test with accommodations and those who took the test under standard conditions. These empirical ICCs were then plotted along with the operational ICCs for these same items. The plots were then sorted into one of the three categories given above.

Table 6.10 — Mathematics accommodation booklet percentages of IRT items by fit categories

Sample	Judged type of IRT curve fit: Nonlogistic	Different Parameters	Fit with Standard Parameters
Grade 4			
Accommodated students	24	55	21
Standard assessment students	10	41	48
	Chi-square = 18.73 significant at the .05 level		
Grade 8			
Accommodated students	11	81	8
Standard assessment students	3	59	38
	Chi-square = 27.59 significant at the .05 level		
Grade 12			
Accommodated students	28	65	8
Standard assessment students	20	58	23
	Chi-square = 8.99 significant at the .05 level		

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table 6.10 presents the percentages of items sorted into each of the three categories for accommodated students and standard assessment students. Also shown are the chi-square tests for homogeneity of proportions. At all three grades, the tests were significant at the .05 level. The numbers of items judged as misfitting nonlogistic were greater for the plots based on accommodated student data than for those based on data from standard assessment students. This result is somewhat at odds with the results of the mathematics DIF analysis of items, where relatively little evidence of differential functioning was found.

A comparison to table 6.8 may provide help in interpreting these data. At the fourth-grade level, the percentages of items in the three categories of fit are about the same as those for science for the accommodated students. For students assessed under standard conditions, on the other hand, there are fewer mathematics items fitting the standard parameters and more mathematics items fitting with different parameters than was the case for science.
For accommodated students at the eighth- and twelfth-grade levels, there are fewer items in

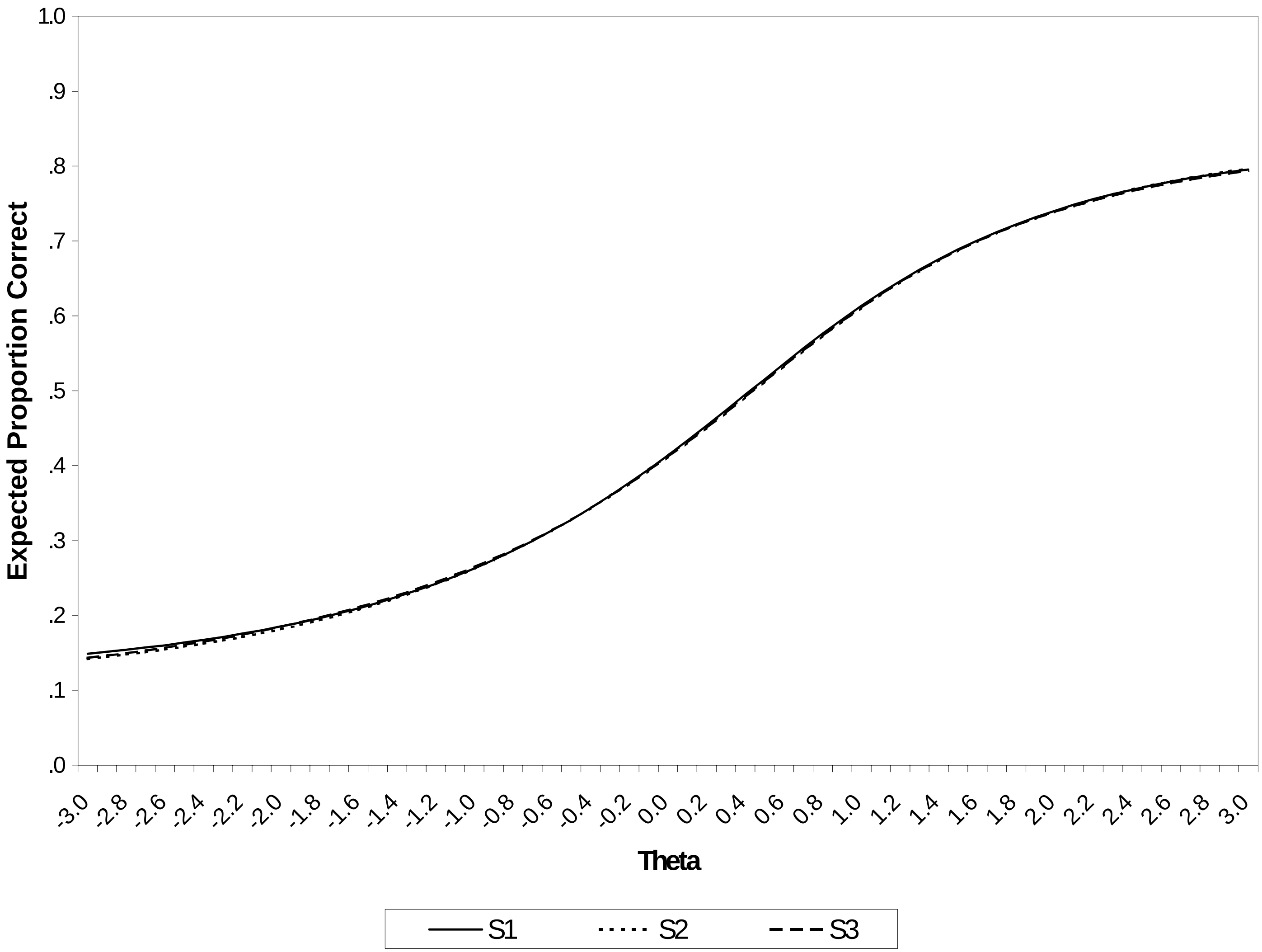
mathematics displaying nonlogistic empirical curves and more appearing to fit with different parameters than in the comparable science data. For the students assessed under standard conditions at eighth grade, there are again fewer items in mathematics displaying nonlogistic curves than in science, but there is a concomitant increase in the category of items that fit the parameters from the standard assessment. The twelfth-grade data display a slightly different change. There are fewer items in mathematics fitting the standard assessment parameters and more that appear to fit a logistic IRT model having different parameters than was the case in science. In summary, although the results show similar trends in mathematics and science, there were more items in mathematics than in science for which data from students assessed under standard conditions fit a model having parameters that were different from the standard parameters. This result may be at least partly a function of mathematics having fewer items per scale than science.

Mathematics Test Characteristic Curves

This section illustrates the similarity of the measurement scales in the three mathematics samples by comparing the test characteristic curves. An example of a test characteristic curve, for the grade-4 numbers and operations scale, is shown in figure 6.10. As in the previously discussed science scales, the proficiency scale in the figure is taken from the scaling program result (mean near zero and standard deviation near one) after the results for the samples were transformed to a common metric, but before transformation to the final 1996 NAEP mathematics reporting metric. Figure 6.10 shows that the test characteristic curve for fourth-grade numbers and operations is indistinguishable for the three different samples.

Figure 6.10 — Test information curves for grade-4 numbers and operations scale

THE NATION'S REPORT CARD naep

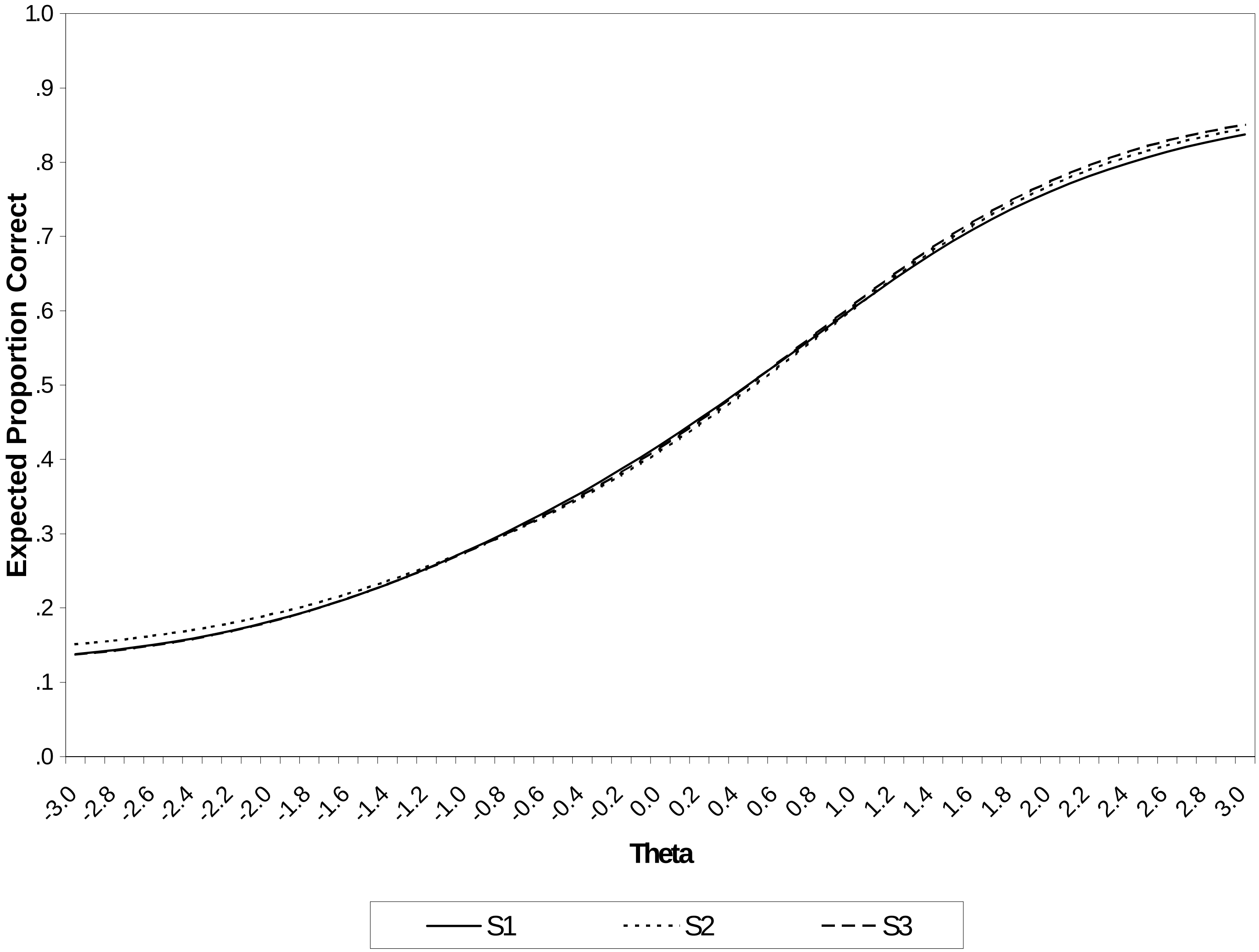


SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figures 6.11 through 6.14 show the test characteristic curves for the other four scales in the fourth grade. The other scales show more differences. This phenomenon probably reflects the numbers of items on the different content strand scales. There are 86 items in numbers and operations at the fourth-grade level, but only 38, 32, 24, and 26 items, for measurement, geometry, data analysis, and algebra and functions, respectively. All other things being equal, fewer items will result in greater variability between test characteristic curves from independent samples.

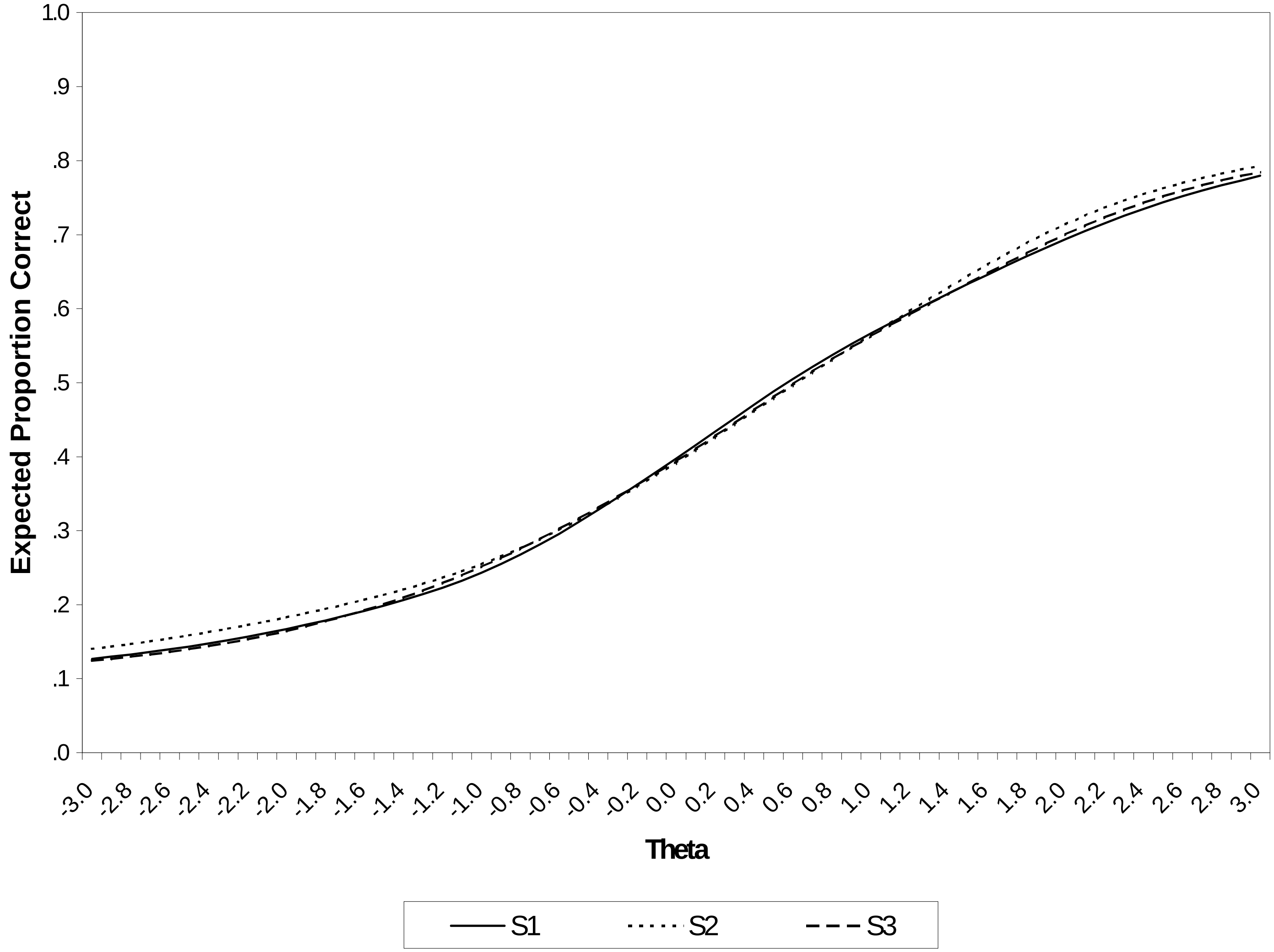
Test characteristic curves for one example each of the eighth- and twelfth-grade scales are shown in figures 6.15 and 6.16. The others are provided in appendix E. The curves shown in these

Figure 6.11 — Test characteristic curves for grade-4 measurement scale



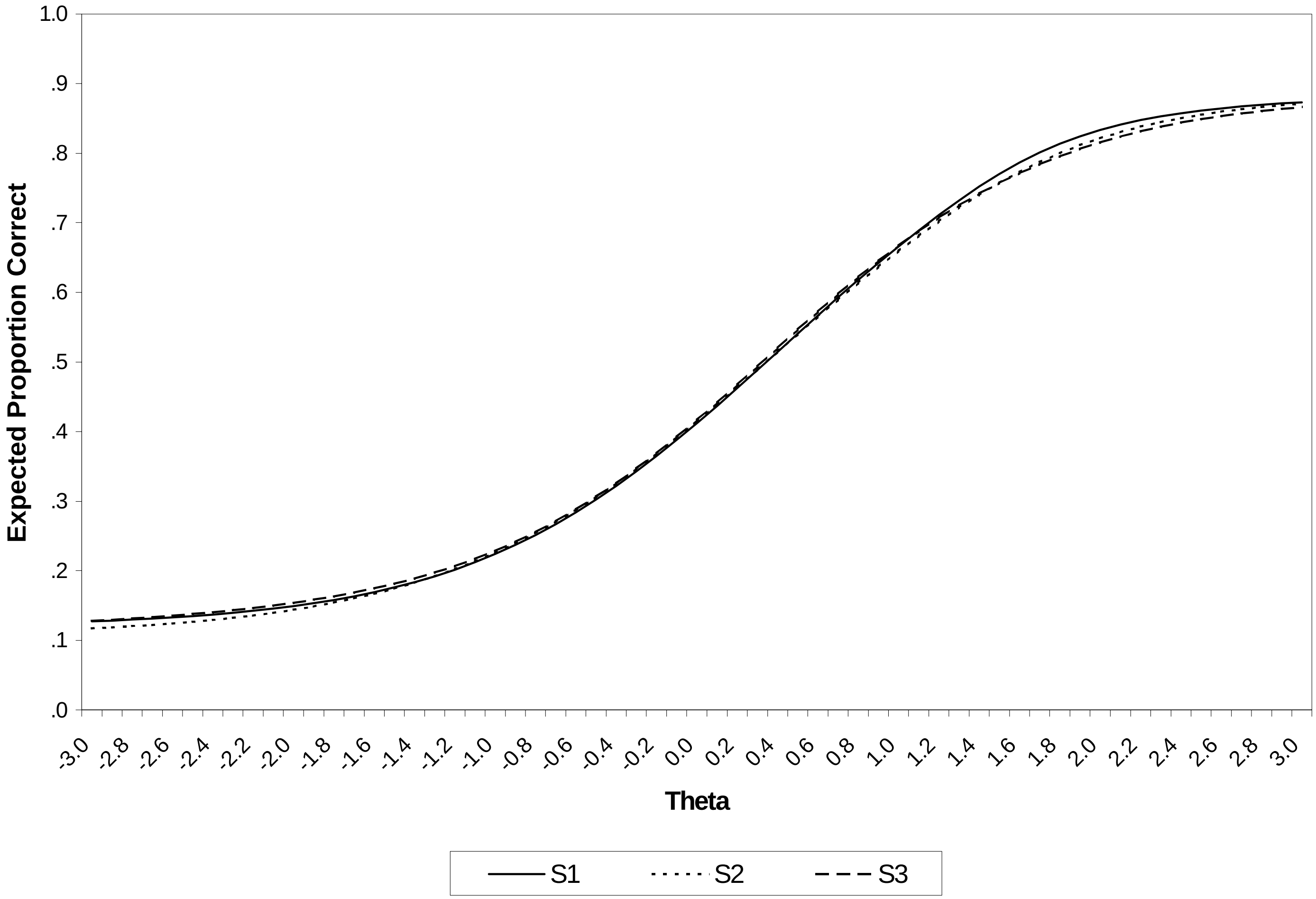
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure 6.12 — Test charateristic curves for grade-4 geometry scale



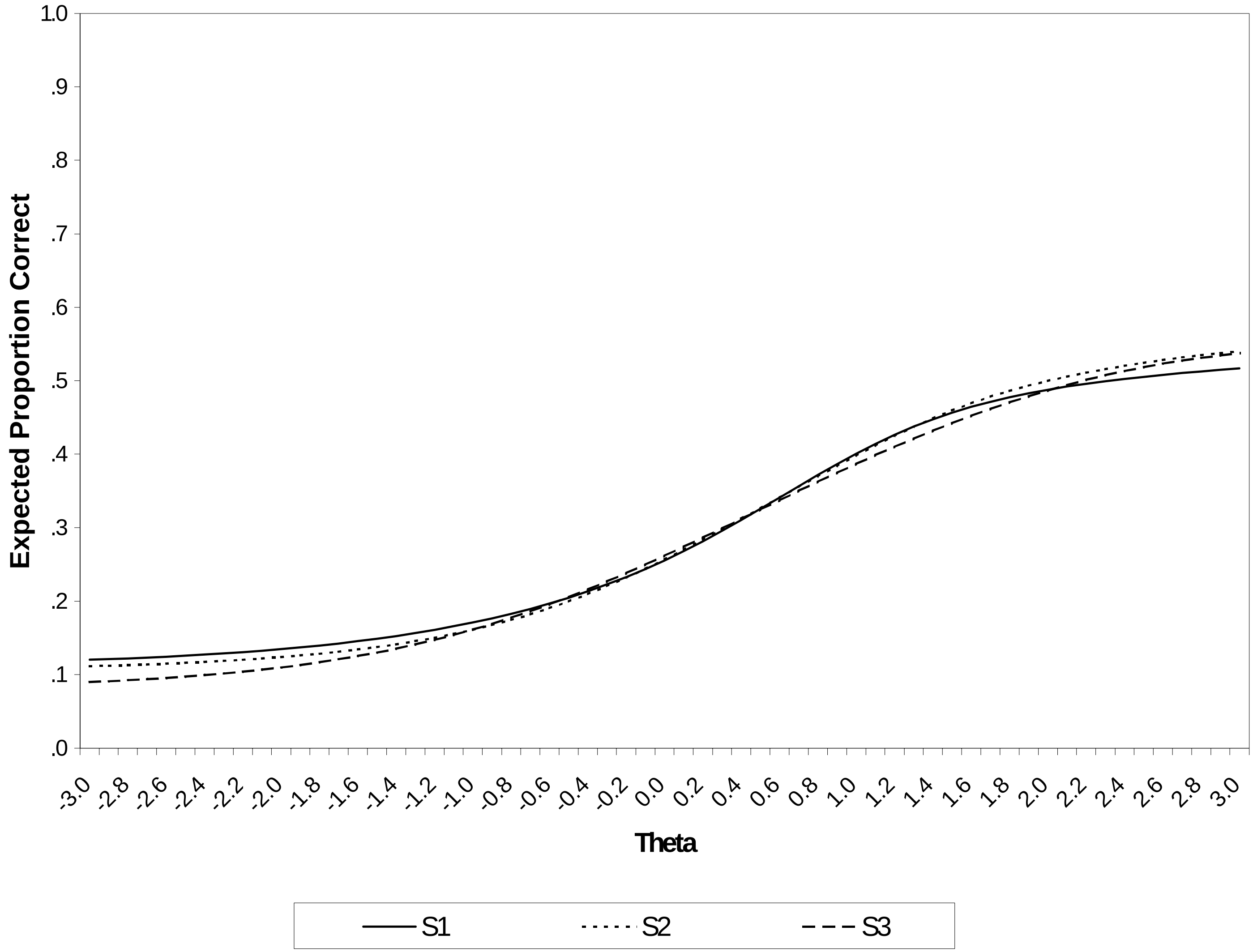
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure 6.13 — Test charateristic curves for grade-4 data analysis scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure 6.14 — Test characteristic curves for grade-4 algebra and functions scale

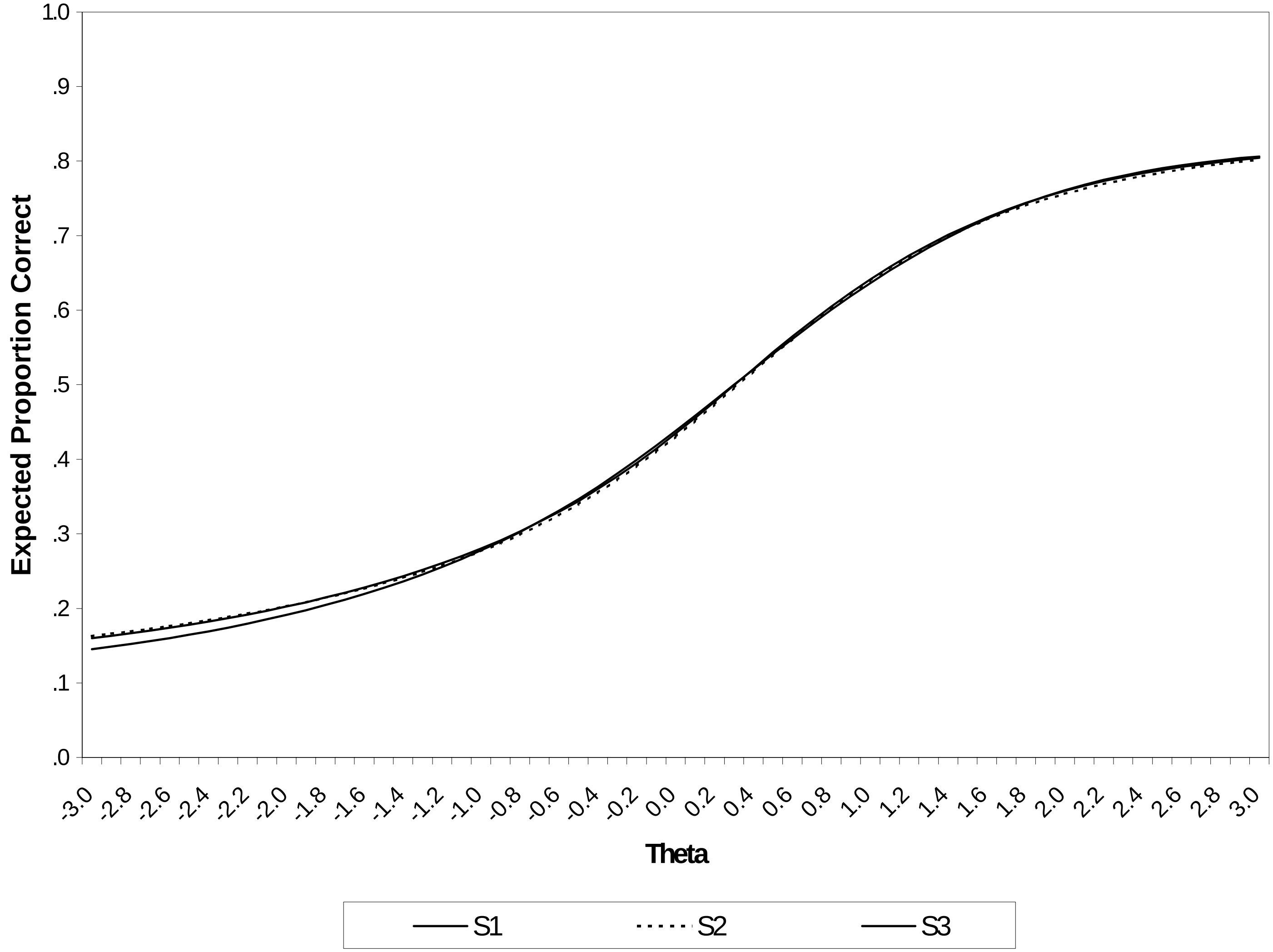


SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

two figures represent the scales at each grade that had the most discrepancy between the curves of the three samples. Note that the twelfth-grade measurement scale (figure 6.16) is the one with the fewest items, 35, of any scale at that grade level. The other content strand scales contained 41, 48, 60, and 62 items, respectively. In the eighth grade, the numbers and operations scale (figure E.1) has the most items and displays the least difference between test characteristic function curves in the three samples. The other four eighth-grade scales, including algebra and functions (figure 6.15), all have about the same number of items. Overall, the curves are quite similar across the samples representing different inclusion criteria and accommodations at all three grade levels. Most of the differences between the curves occur in the tails of the distributions. Typically, any test measures less precisely in the tails of the distribution, where few examinees scores may be expected to fall, than in the central range where the bulk of students' scores may be expected to be observed. Hence, one would expect to observe more variability in the tails of curves for independent samples, such as those plotted here.

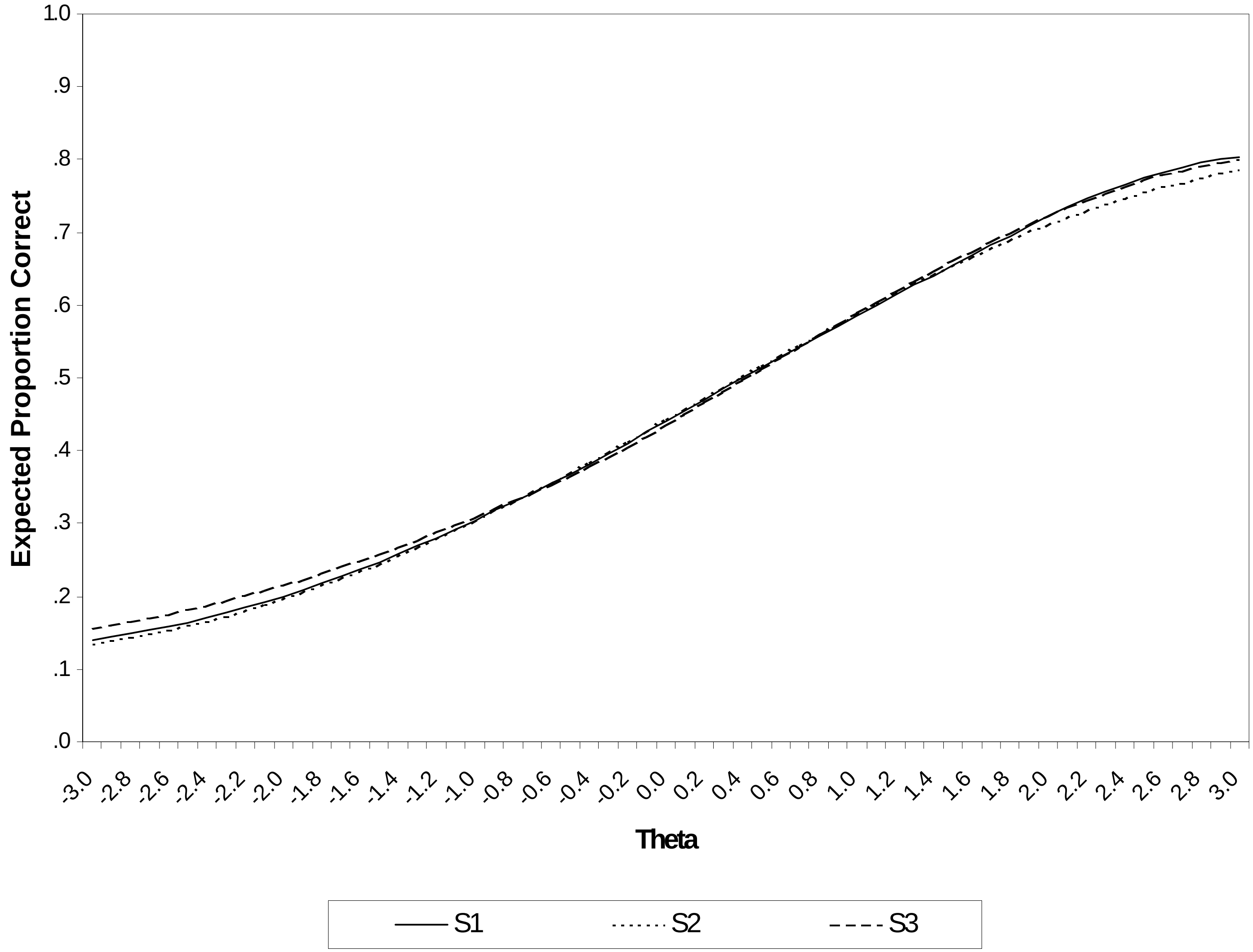
Figure 6.15 — Test characteristic curves for grade-8 algebra and functions scale

THE NATION'S REPORT CARD naep



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure 6.16 — Test characteristic curves for grade-12 measurement scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Mathematics Test Information Curves

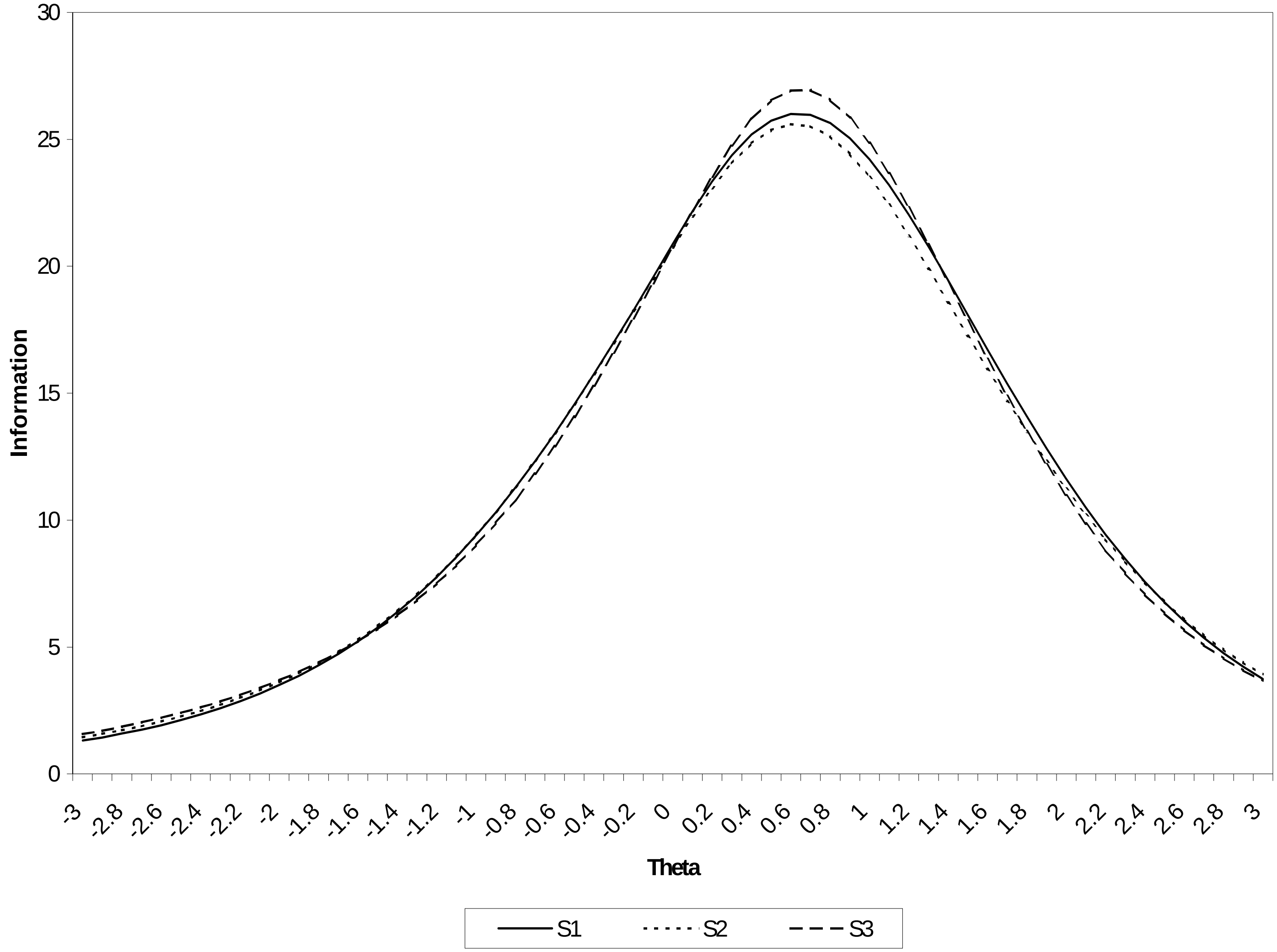
Figure 6.17 shows one set of test information plots for the three different fourth-grade mathematics samples — for the numbers and operations scale. As mentioned previously, that scale has many more items than do the other fourth-grade scales. For this scale, the test information curves are remarkably similar across the three samples. The curves for the other four content strand scales at this grade level are shown in figures 6.18 through 6.21. The larger differences in the curves from the different samples for some scales (for example algebra and functions showing less spread for S1 than for S2 and S3) probably results from estimation with relatively small numbers of items per scale.

Figures in appendix F show similar curves for the eighth- and twelfth-grade mathematics scales. The information patterns tend to be very similar across the three samples in all these figures.

The reader may have noted that the mathematics information curves display larger differences than those for science. In interpreting the information function plots, the reader is reminded that they represent data about the precision of IRT estimation of students' hypothetical proficiencies. As mentioned previously in the section on science results, such estimates are highly variable from sample to sample for scales based on small numbers of items administered to each student. Each of the five NAEP mathematics scales (at each grade level), in general, is based on fewer items than each of the three science scales. Hence, the information curves are likely to show more sample-to-sample fluctuation in mathematics.

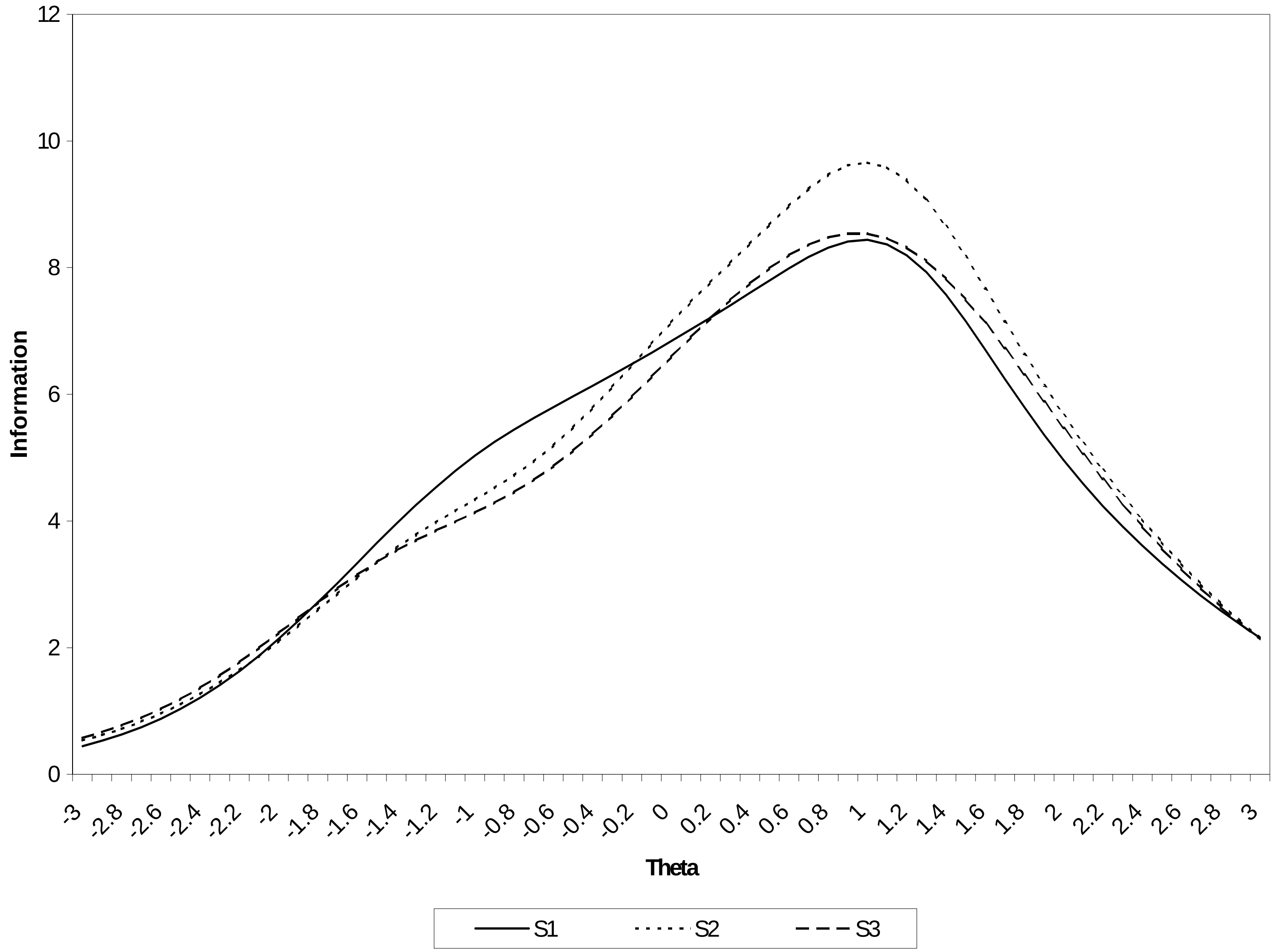
THE NATION'S REPORT CARD naep

Figure 6.17 — Test information curves for grade-4 numbers and operations scale



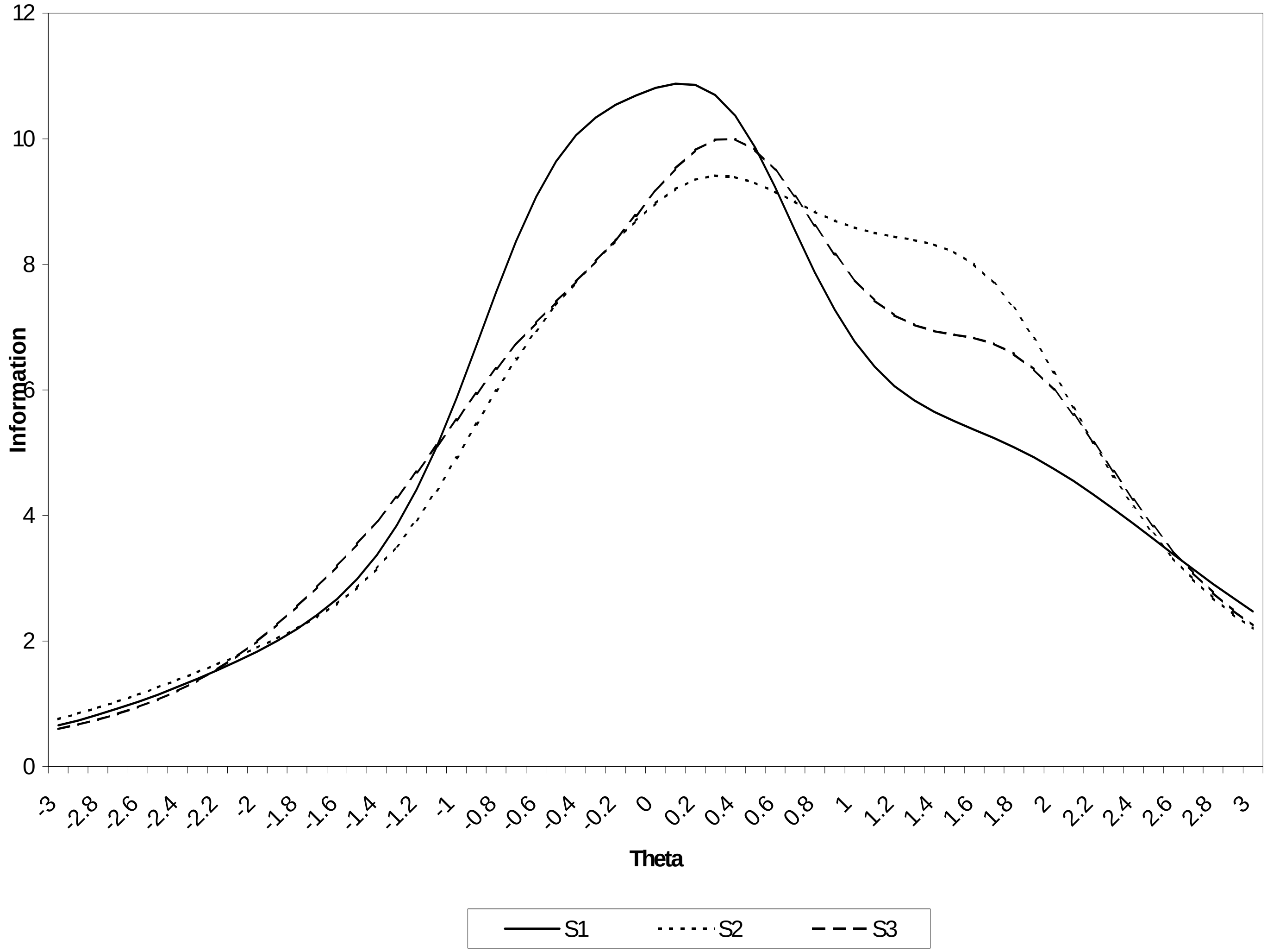
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure 6.18 — Test information curves for grade-4 measurement scale



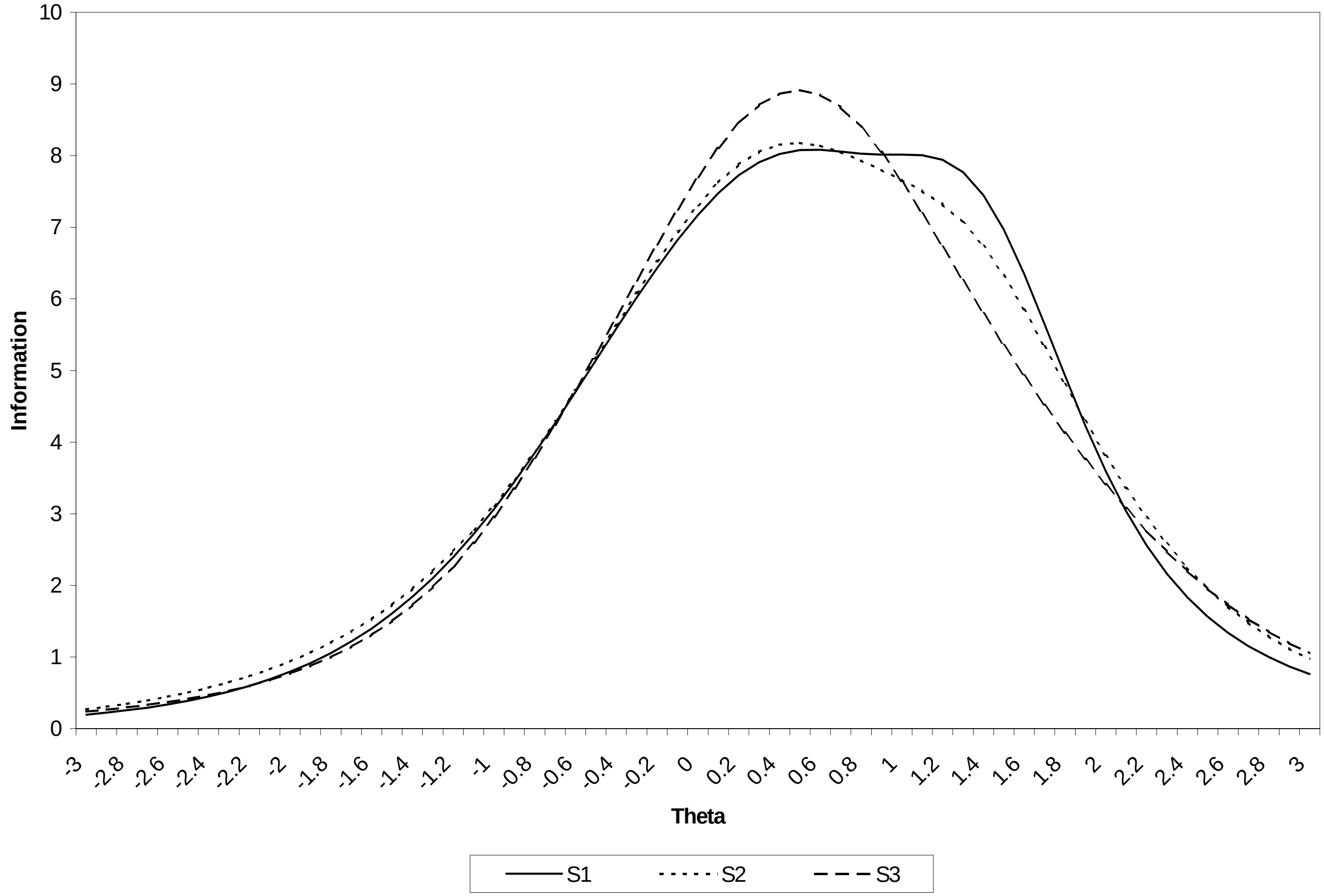
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure 6.19 — Test information curves for grade-4 geometry scale



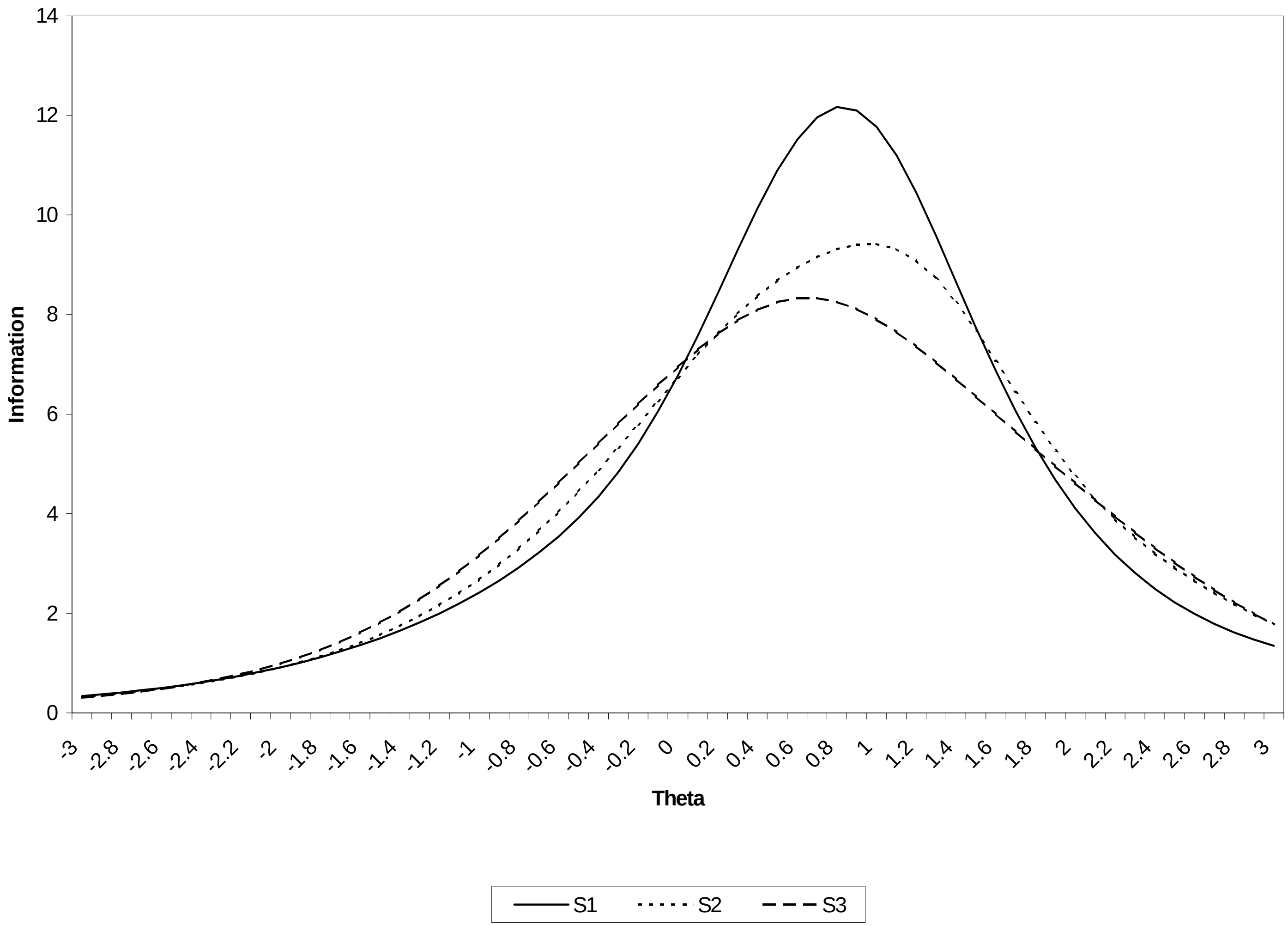
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure 6.20 — Test information curves for grade-4 data analysis scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure 6.21 — Test information curves for grade-4 algebra & functions scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Chapter Summary

Taking into account all of the results presented in this chapter, a reasonable conclusion is that the NAEP scales in mathematics and science are neither greatly affected by changes in the inclusion criteria, nor by the presence of accommodations for students with disabilities and limited English proficient students.

The science data provide the best baseline for comparisons because one sample was sufficiently large to allow division into two random halves and enabled examination of random sample-to-sample fluctuations in various statistics. The difference in results between these samples can be examined in the light of the foregoing baseline analyses. The findings suggest that the ***differences attributable to differences in administration procedures in the SD/LEP student subpopulations appear to be no more than would be expected to occur in repeated random sampling under identical administration procedures.***

At two of the three grades in science there is some evidence of accommodations-related differential item functioning (DIF), generated with relatively small sample sizes. In contrast, the mathematics assessment provided little, if any, evidence of accommodations-related DIF. Some differences were observed in the curves of fit of item response data to the theoretical response curves for both science and mathematics at all grade levels. However, it cannot be concluded with certainty that these results are not attributable to the variation expected with such small numbers of students whose data were used to construct these curves.

The most compelling evidence is probably exhibited in the test characteristic curves. They show that the change in inclusion criteria and the inclusion of accommodated students had essentially no effect on the scales derived for reporting the results for the overall NAEP samples of students.

Chapter 7

Comparison of Scale Score Means

Introduction

One of the goals of this report was to determine whether the proposed changes to NAEP inclusion criteria and the provision of accommodations and adaptations would affect estimated distributions of scale scores for those NAEP reporting groups. The proposed procedural changes were intended to affect the populations to which NAEP results apply, specifically by increasing the numbers of special needs students. Moreover, the data from which NAEP results are estimated would include data from students tested with accommodations and adaptations. The question addressed in this chapter is whether these changes would alter basic NAEP results. The production of such changes as a result of methodological revisions poses threats to the validity of comparisons of current NAEP results to those from previous assessments.

One of the typical ways in which NAEP data are reported is by the mean scale score of important reporting groups of students. Besides the total group at a given grade level, the main reporting groups are based on variables such as gender, race/ethnicity, and parents' education. This chapter presents the results of using the standard NAEP analysis model, described in the *1996 NAEP Technical Report*,[1] to estimate the mean scale score results for the reporting subgroups of the national samples. Two main sections contain separate results from the two subject areas. Subsections of this chapter contain descriptions of the methodology and comparisons of the subgroup means.

The focus of this chapter is on differences in groups defined by the reporting variables because the main purpose of this study is to determine whether including accommodated students in the NAEP analyses significantly changes scale score results. A few differences between specific groups that are not reporting groups in the standard reports are also pointed out here because they may be of interest to readers of this report.

---

[1] Allen, N.L., Carlson, J.E., & Zelenak, C.A. (forthcoming). *The NAEP 1996 technical report.* Washington, DC: National Center for Education Statistics.

General Procedures

The NAEP procedures for estimating scale score results make use of statistical models that use the students' responses to each test item, the item response theory parameter estimates for each item, student demographic information, and student, teacher, and school background questionnaire responses. Using the standard NAEP procedures, separate estimations were carried out for each sample for each subject area at each grade level. This required four analyses at each grade level for science (S2, S3, and the two random halves of S2) and three analyses at each grade level for mathematics (S1, S2, and S3). The item parameter estimates that were used were those computed for the results described in chapter 6. Unlike the item scaling in chapter 6, the scale score estimation models used here are multivariate ones, with results being estimated simultaneously for the three field-of-science scales or the five mathematics content strand scales.

Standard NAEP estimation procedures were used to compute data summary tables, the results of which are discussed in the remaining sections of this chapter. Mean scale scores were computed for subgroups based on the main reporting variables, as described below. In addition, tests of statistical significance were conducted to support inferences about mean differences between S1, S2, and S3 for mathematics and between S2 and S3 for science, as well as for science's two random half samples from S2. These tests were carried out at the .05 level of significance, within appropriate adjustment for multiple comparison based on the Bonferonni procedure.

The following are the main reporting variables:

- **Gender:** male, female;
- **Race/Ethnicity:** Black, American Indian, Asian/Pacific Islander, Hispanic, White;
- **Parents' highest level of education:** less than high school, graduated high school, some education after high school, graduated college, unknown.
- **Type of school:** public, nonpublic, private (non-Catholic) only, Catholic only. (The nonpublic classification includes the last two, so this classification does not result in four independent subsamples);
- **Region of the country:** Northeast, Southeast, Central, West;
- **Type of location:** central city, urban fringe/large town, rural/small town;
- **Title I participation:** participated, did not participate; and
- **National school lunch program:** eligible, not eligible.

Further details on the formation of these reporting variables are provided in the report cards for the 1996 science and mathematics assessments[2].

In addition to the above classifications, breakdowns were made by SD/LEP classifications for making comparisons between the various subsamples based on inclusion rules and accommodations provided. These are:

- **SD/LEP status:** SD and/or LEP, neither SD nor LEP;
- **SD status:** SD, not SD;
- **LEP status:** LEP, not LEP; and
- **Combined status:** SD and LEP, SD only, LEP only.

The standard NAEP procedure is to report only those results for groups that are sufficiently large that reliable standard errors of the means can be computed. In a few cases, a subgroup contained too few students for reliable reporting or valid significance testing. This occurred most frequently for the American Indian subgroup. Some results for the standard NAEP reporting groups may be of interest to readers, so they are reported in this chapter.[3]

Science Results

Specific procedures used with the science assessment data and certain scale score results are reported in this section. Parallel results for mathematics are reported in the next section.

Science Procedures. Scale score estimation was carried out within each of the four samples (S2, S3, and the two half samples of S2) in each of the three grade levels. As indicated above, because S2 was the national reporting sample, it was not necessary to perform a new analysis on this sample. Therefore, only nine new analyses needed to be performed.

Results for each analysis were transformed to the final NAEP science reporting metric. The 1996 NAEP science assessment was independently scaled within each grade level. The overall mean for each grade was set to 150 and the standard deviation to 35 on each of the three fields-of-science scales and a composite of those scale scores.

Science Results for S2 versus S3 as compared to Random Half Samples. A total of 444 statistical comparisons were made between the two random half samples for the subgroups

---

[2] Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). *NAEP 1996 mathematics report card for the nation and states*. Washington, DC: National Center for Education Statistics.

O'Sullivan, C.Y., Reese, C.M., & Mazzeo, J. (1997). *NAEP 1996 science report card for the nation and states.* Washington, DC: National Center for Education Statistics

[3] Differences cited between groups are statistically significant. The significance testing procedures used the .05 significance level. The Bonferroni adjustment was incorporated to control for the number of groups being compared, called "family size." For example, the family size for gender comparisons between S2 and S3 is two, because comparisons are made separately for female and male students. In comparisons between pairs of samples within gender the Bonferroni procedure controls the chances of making at least one Type I error (wrongly concluding true differences) at the .05 level when making the two comparisons. Further details about NAEP significance testing procedures are provided in: Allen, N.L., Carlson, J.E., & Zelenak, C.A. (Eds.). (in press). *The NAEP 1996 Technical Report*. Washington, DC: National Center for Education Statistics.

defined by the main reporting variables listed in the General Procedures section across the three grade levels. Of these, nine resulted in statistically significant differences. There are, of course, no true mean differences between random samples. Five percent (or 22) would be expected to be statistically significant at the .05 level if all the tests were between independent samples. The small number of differences that actually were significant is partially a function of use of the Bonferroni procedure to control the chances of making at least one false declaration of significance in each family of comparisons, as described above. It is also partly due to the fact that families of comparisons themselves are not independent.

Out of a similar number of comparisons made on the two samples defined by presence or absence of accommodations (S2 versus S3) only five were significant. In other words, the number of differences that would be concluded to be true differences rather than being attributable to random factors in sampling was actually lower in the two samples for which different administration procedures (accommodations or no accommodations) were used than between two randomly formed subsamples.

The five significant differences in mean scale scores between S2 and S3 all involved limited English proficient (LEP) students, as shown in table 7.1. In each case, the average score was higher for the LEP students in S2 than in S3. The multiple significant results at eighth grade are not independent. First the earth science scale is part of the composite hence the significant results on the two scales are related. Moreover, the three groups involved (all LEP students, LEP students who are not SD, and, combined LEP and SD students) are not independent since the last group subsumes the first two.

On initial examination, the finding of significant differences in average scale scores for groups involving S2 and S3 LEP students would appear to make sense. Chapter 5 results on inclusion for LEP students in the mathematics assessement suggested that the provision of accommodations and adaptations did increase the percentages of LEP students participating in NAEP. If the increases were occurring among lower-scoring LEP students, one might expect the result evident in table 7.1 (i.e., LEP students in S3 have lower average scores than their counterparts in S2). However, as discussed in Chapter 5 and reported in the *NAEP 1996 Science Report Card*, inclusion rates in S2 and S3 for the science assessment did not differ significantly at any of the three grades. Moreover, most of the differences in scale score means reported above are observed at grade 8, where the inclusion rates appeared comparable across the two samples. In light of the science inclusion rate results, the significant differences reported in table 7.1 are puzzling.

The results in table 7.1 should be interpreted with some caution for two reasons. First, results for these particular subgroups are not routinely reported in NAEP. Second, four of the five significant differences involve at least one group average that would not normally be reported in NAEP because the nature of the samples does not permit accurate determination of the sampling variability of the estimates. Using standard NAEP analysis procedures, results are not reported if the coefficient of variation for the weighted sample size used in estimating the standard error of a statistic is such that instability in the estimate is likely. Johnson and Rust[4] provide details of the rationale underlying this judgment. For this report these results are displayed, but with a qualifying footnote.

---

[4] Johnson, E.G., & Rust, K. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics,* 17, 175-190.

Table 7.1 — Means with significant differences between samples S2 and S3: Science, grades 4 and 8

	Students from sample S2		Students from sample S3	
	N	Mean	N	Mean
LEP Students (some are also SD)				
Grade 4/physical science	261	113 !	174	98 ††!
Grade 8/composite	217	107	182	94 ††!
Grade 8/earth science	217	113	182	93 ††!
LEP Students (Not also SD)				
Grade 8/earth science	203	113	167	97 ††!
Combined SD and LEP Students				
Grade 8/earth science	652	123	527	112 ††

†† Indicates a significant difference between S2 and S3 results.

! Interpret with caution: the nature of the sample does not allow accurate determination of the variability of this statistic.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Tables 7.2 through 7.7 present average scale scores by sample type for two of NAEP's traditional reporting groups, gender and race/ethnicity. Tables 7.2, 7.3, and 7.4 present the results by gender. At all three grades, little difference was evident in the S2 and S3 means for males and females and, as noted earlier, none of the differences are statistically significant. This result is not surprising in that the gender subgroups are large and it is unlikely that their results would be noticeably impacted by changes in the results for SD or LEP students. The latter groups contribute a relatively small proportion of the data involved in the calculation of the gender means. It should be noted that far more males than females are classified as SD. Thus, if an effect were present it would most likely be evident in the results for males. Across the three grades there is no apparent pattern of larger differences by sample type for males. Tables 7.5, 7.6 and 7.7 present the results by race/ethnicity. Again, little difference is evident in the means of these groups across the sample types, and none of these apparent differences is statistically significant. The lack of noticeable impact in the race/ethnicity group results is a bit more surprising. Black students are disproportionately represented among students with disabilities and, particularly at the lower two grades, Hispanic students are a large percentage of LEP students. Despite these facts, the results for Black students show virtually no apparent difference by sample type . The apparent differences for Hispanic students follow differing patterns across the grades.

As noted in table 7.1, average scores were lower in S3 than in S2 on some scales for certain LEP groups. At grade 4, this result did not appear to track through into the average scores for hispanic students. In fact, the average scores in S3 for Hispanic students at grade 4 (table 7.5) and grade 12 (table 7.7) appear slightly higher than those in S2 on virtually all of the scales. In contrast, there is some evidence that the lower average scores for the LEP groups had an impact on the average scores for Hispanic students at grade 8. The average scores for grade-8 Hispanic students appeared lower in S3 than in S2 for all scales. It is important to note, however, that none of these apparent differences is statistically significant.

Table 7.2 — Grade-4 science means overall and by gender

	Students from sample S2		Students from sample S3	
	N	Mean	N	Mean
Composite	**7,305**	**150**	**4,273**	**150**
Female	**3,654**	**149**	**2,126**	**149**
Male	**3,651**	**151**	**2,147**	**151**
Physical science		150		150
Female		150		150
Male		150		150
Earth science		150		150
Female		147		148
Male		153		152
Life science		150		150
Female		150		150
Male		150		150

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

THE NATION'S REPORT CARD naep

Table 7.3 — Grade-8 science means overall and by gender

	Students from sample S2		Students from sample S3	
	N	Mean	N	Mean
Composite	**7,774**	**150**	**4,197**	**150**
Female	**3,902**	**149**	**2,099**	**149**
Male	**3,872**	**151**	**2,098**	**151**
Physical science		150		150
Female		148		149
Male		152		151
Earth science		150		150
Female		148		148
Male		152		152
Life science		150		150
Female		151		150
Male		149		150

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

THE NATION'S REPORT CARD naep

Table 7.4 — Grade-12 science means overall and by gender

	Students from sample S2		Students from sample S3	
	N	Mean	N	Mean
Composite	**7,537**	**150**	**3,943**	**150**
Female	**3,990**	**148**	**1,932**	**146**
Male	**3,547**	**152**	**2,011**	**154**
Physical science		150		150
Female		147		146
Male		153		154
Earth science		150		150
Female		146		145
Male		154		155
Life science		150		150
Female		150		149
Male		150		152

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table 7.5 — Grade-4 science means by race/ethnicity

	Students from sample S2		Students from sample S3	
	N	Mean	N	Mean
Composite				
Black	**1,251**	**124**	**740**	**123**
American Indian	**223**	**144**	**160**	**140**
Asian/Pacific Islander	**356**	**151**	**188**	**152**
Hispanic	**1,352**	**128**	**818**	**130**
White	**4,106**	**160**	**2,350**	**160**
Physical science				
Black		123		122
American Indian		145		138
Asian/Pacific Islander		150		146
Hispanic		126		128
White		160		160
Earth science				
Black		124		122
American Indian		144		141
Asian/Pacific Islander		153		156
Hispanic		129		132
White		159		159
Life science				
Black		124		123
American Indian		144		142
Asian/Pacific Islander		150		155
Hispanic		129		130
White		159		159

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table 7.6 — Grade-8 science means by race/ethnicity

	Students from sample S2		Students from sample S3	
	N	Mean	N	Mean
Composite				
Black	**1,492**	**121**	**804**	**122**
American Indian	**149**	**148**	**64**	**137**
Asian/Pacific Islander	**382**	**152**	**182**	**153**
Hispanic	**1,426**	**129**	**726**	**125**
White	**4,292**	**160**	**2,411**	**160**
Physical science				
Black		119		120
American Indian		149		136
Asian/Pacific Islander		149		154
Hispanic		129		126
White		160		160
Earth science				
Black		119		122
American Indian		148		139
Asian/Pacific Islander		155		153
Hispanic		130		126
White		160		160
Life science				
Black		124		124
American Indian		148		137
Asian/Pacific Islander		152		152
Hispanic		128		124
White		159		160

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table 7.7 — Grade-12 science means by race/ethnicity

	Students from sample S2		Students from sample S3	
	N	Mean	N	Mean
Composite				
Black	**1,225**	**124**	**685**	**123**
American Indian	**70**	**145**	**38**	**——**
Asian/Pacific Islander	**458**	**149**	**237**	**146**
Hispanic	**1,015**	**130**	**554**	**133**
White	**4,748**	**159**	**2,427**	**158**
Physical science				
Black		124		122
American Indian		145		——
Asian/Pacific Islander		151		148
Hispanic		130		132
White		158		159
Earth science				
Black		122		123
American Indian		148		——
Asian/Pacific Islander		147		143
Hispanic		130		135
White		159		158
Life science				
Black		125		125
American Indian		142		——
Asian/Pacific Islander		149		146
Hispanic		129		132
White		159		158

NOTE: —— Indicates insufficient sample size to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Mathematics Results

This section describes the procedures specific to the mathematics data and presents results for the 1996 mathematics assessment, similar to those for science presented above.

Mathematics Procedures. Scale score estimation was carried out separately within each of the three samples in each of the three grades, so that nine separate analyses were performed. The 1996 NAEP mathematics assessment was scaled to the same metric as the 1990 and 1992 assessments in order that trends could be studied. In this study, the data were scaled to this same metric. Scaling was performed independently within each grade level, but results were linked to the cross-grade scale established in 1990 using the same procedures as the main assessment. Because the results reported in this section were placed on the previously defined NAEP mathematics reporting scale, in contrast to the science results, the mathematics results reflect different means for the three grade levels.

Mathematics Scale Score Results. A total of 1,332 comparisons were made between reporting group scale score means in the S1, S2, and S3 samples. As noted earlier, tests of significance were carried out at the .05 level of significance, with appropriate controls for multiple comparisons based on a Bonferroni procedure. Across all three grades, only three significant differences emerged from these comparisons. The number of significant differences found was in fact less than the number encountered for the two random half-samples of the science S2 samples reported above. Thus, as was the case for the science samples, the differences in scale score results for the different sample types (as defined by the differing inclusion methodologies) are no greater than those that would be expected on the basis of sampling variability alone.

The three significant differences that were found are given in table 7.8. All three were at grade 12. The overall average for the geometry scale, and the average for female students, was lower in the S3 sample than in the S1 sample. These two results are clearly dependent, as females make up nearly half of the total result. The third significant difference involved the algebra scale where the average score was higher for Asian/Pacific Islander students in S1 than in S2.

Table 7.8 — Mathematics significant differences

	Students from sample S1		Students from sample S2		Students from sample S3	
	N	Mean	N	Mean	N	Mean
All Students						
Grade 12/geometry	3,616	308	3,607	305	3,437	304†
Female Students						
Grade 12/geometry	1,888	307	1,900	303	1,688	301†
Asian/Pacific Islander Students						
Grade 12/algebra	181	336	200	314‡	243	318

‡ Indicates a significant difference between S1 and S2 results.

† Indicates a significant difference between S1 and S3 results.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

There is little justification to believe that these three differences are related to the changes in the inclusion procedures under study. First, as reported in chapters 2 through 5, almost no differences betweeen the samples on mathematics inclusion rates were observed at grade 12. Moreover, as discussed with respect to the science results, there is little reason to expect a result for females at any grade, since it is males, not females, that are overrepresented in the SD population. In short, given the small number of significant differences and the absence of a sensible explanation, it is more reasonable to assume these differences are simply "type I errors."

Tables 7.9 through 7.14 present average scale scores by sample type for the gender and race/ethnicity reporting groups. Tables 7.9, 7.10, and 7.11 present the results by gender. At the two lower grades, few differences and little pattern is evident in gender means by sample type and, as noted earlier, none of the differences is statistically significant. Given that the provision of accommodations and adaptations did result in increased inclusion at these grades, it is reasonable to expect that if an effect were to be noted it would be at these grades. Thus, the absence of any evidence of impact is noteworthy. The results at grade 12 reflect the differences in S1 and S3 geometry scale score averages discussed earlier. As noted above, there is little reason to believe these differences are related to the inclusion policies under study.

Tables 7.12, 7.13 and 7.14 present the results by race/ethnicity. Again, little difference is evident in the means of these groups across the sample types and all but one of these apparent differences (the aforementioned difference in algebra scale scores between S1 and S2 for Asian students) are not statistically significant. At all three grades, the average scores for Black students in S3 appear slightly lower than those in S1 and S2. Such a pattern is plausible given the disproportionate representation of Black students among those with disabilities. However, none of these differences is statistically significant. At grade 12, average scores in S2 and S3 (the revised inclusion criteria) also appear lower than the corresponding S1 (original criteria) for Asian students and Hispanic students. However, only the former is statistically significant. Moreover, there was little in the inclusion rate data to suggest that the changed policies were resulting in any impact at this grade. Thus, in total there is little evidence to suggest that mathematics average scale scores were affected by the changes in inclusion policies for any of the gender or race/ethnicity groups.

Table 7.9 — Grade-4 mathematics means overall and by gender

THE NATION'S REPORT CARD naep

	Students from sample S1		Students from sample S2		Students from sample S3	
	N	Mean	N	Mean	N	Mean
Composite	**3,604**	**224**	**3,334**	**225**	**3,892**	**223**
Female	**1,773**	**222**	**1,684**	**224**	**1,941**	**223**
Male	**1,831**	**226**	**1,650**	**225**	**1,951**	**223**
Numeric operations		221		222		220
Female		219		222		220
Male		223		223		220
Measurement		225		227		225
Female		222		226		224
Male		228		229		225
Geometry		225		225		223
Female		224		226		225
Male		226		224		222
Data analysis		225		225		224
Female		222		226		224
Male		227		225		224
Algebra		228		228		228
Female		224		227		228
Male		231		230		227

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

THE NATION'S REPORT CARD naep

Table 7.10 — Grade-8 mathematics means overall and by gender

	Students from sample S1		Students from sample S2		Students from sample S3	
	N	Mean	N	Mean	N	Mean
Composite	**3,941**	**273**	**3,671**	**270**	**3,909**	**272**
Female	**1,970**	**273**	**1,776**	**270**	**1,928**	**271**
Male	**1,971**	**273**	**1,895**	**271**	**1,981**	**273**
Numeric operations		275		272		274
Female		275		272		273
Male		275		272		276
Measurement		272		268		270
Female		269		265		266
Male		274		270		274
Geometry		270		269		270
Female		270		268		269
Male		269		269		270
Data analysis		274		271		272
Female		276		271		271
Male		273		271		273
Algebra		274		271		272
Female		275		270		272
Male		274		272		273

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table 7.11 — Grade-12 mathematics means overall and by gender

	Students from sample S1		Students from sample S2		Students from sample S3	
	N	Mean	N	Mean	N	Mean
Composite	**3,616**	**305**	**3,607**	**303**	**3,437**	**302**
Female	**1,888**	**304**	**1,900**	**301**	**1,688**	**300**
Male	**1,728**	**306**	**1,707**	**304**	**1,749**	**303**
Numeric operations		302		300		299
Female		300		299		298
Male		303		302		301
Measurement		303		301		301
Female		301		297		298
Male		306		305		304
Geometry		308		305		304 †
Female		307		303		301 †
Male		310		308		307
Data analysis		305		303		301
Female		304		303		302
Male		305		303		301
Algebra		306		303		303
Female		307		303		303
Male		306		303		304

† Indicates a significant difference between S1 and S3 results.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table 7.12 — Grade-4 mathematics means by race/ethnicity

	Students from sample S1		Students from sample S2		Students from sample S3	
	N	Mean	N	Mean	N	Mean
Composite						
Black	**574**	**201**	**555**	**200**	**663**	**196**
American Indian	**94**	**216**	**65**	**214**	**122**	**212**
Asian/Pacific Islander	**143**	**231**	**139**	**236**	**188**	**235**
Hispanic	**508**	**207**	**585**	**206**	**714**	**206**
White	**2,268**	**232**	**1,983**	**233**	**2,193**	**231**
Numeric operations						
Black		199		199		193
American Indian		213		208		208
Asian/Pacific Islander		228		236		233
Hispanic		203		203		204
White		229		230		229
Measurement						
Black		199		201		192
American Indian		217		219		215
Asian/Pacific Islander		232		231		235
Hispanic		209		207		204
White		234		236		235
Geometry						
Black		203		199		201
American Indian		220		220		214
Asian/Pacific Islander		235		240		240
Hispanic		208		209		210
White		233		233		230
Data analysis						
Black		198		197		196
American Indian		217		211		211
Asian/Pacific Islander		230		234		236
Hispanic		208		208		205
White		234		234		233
Algebra						
Black		206		206		204
American Indian		221		219		217
Asian/Pacific Islander		234		239		239
Hispanic		214		211		214
White		235		236		235

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table 7.13 — Grade-8 mathematics means by race/ethnicity

	Students from Sample S1		Students from Sample S2		Students from Sample S3	
	N	Mean	N	Mean	N	Mean
Composite						
Black	**530**	**244**	**716**	**241**	**652**	**239**
American Indian	**73**	**263 !**	**42**	**— —**	**40**	**— —**
Asian/Pacific Islander	**265**	**273**	**180**	**279**	**171**	**280**
Hispanic	**397**	**251**	**698**	**253**	**648**	**251**
White	**2,658**	**283**	**2,030**	**279**	**2,389**	**282**
Numeric operations						
Black		249		248		246
American Indian		269 !		— —		— —
Asian/Pacific Islander		277		277		284
Hispanic		254		256		253
White		284		280		284
Measurement						
Black		229		228		223
American Indian		258 !		— —		— —
Asian/Pacific Islander		269		280		277
Hispanic		245		245		244
White		285		280		284
Geometry						
Black		242		241		241
American Indian		262 !		— —		— —
Asian/Pacific Islander		270		278		276
Hispanic		252		256		255
White		279		276		278
Data analysis						
Black		245		235		237
American Indian		258 !		— —		— —
Asian/Pacific Islander		266		273		277
Hispanic		243		250		248
White		287		282		283
Algebra						
Black		249		246		242
American Indian		265 !		— —		— —
Asian/Pacific Islander		278		283		282
Hispanic		256		253		253
White		283		279		282

— — Indicates insufficient sample size to permit a reliable estimate.
! Interpret with caution: the nature of the sample does not allow accurate determination of the variability of this statistic.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table 7.14 — Grade-12 mathematics means by race/ethnicity

	Students from sample S1		Students from sample S2		Students from sample S3	
	N	Mean	N	Mean	N	Mean
Composite						
Black	**548**	**282**	**591**	**278**	**630**	**276**
American Indian	**90**	**279 !**	**32**	**— —**	**25**	**— —**
Asian/Pacific Islander	**181**	**326**	**200**	**311**	**243**	**310**
Hispanic	**300**	**289**	**545**	**284**	**532**	**285**
White	**2,490**	**312**	**2,230**	**310**	**2,007**	**310**
Numeric operations						
Black		281		280		275
American Indian		275 !		— —		— —
Asian/Pacific Islander		320		311		305
Hispanic		286		282		282
White		308		307		307
Measurement						
Black		273		272		268
American Indian		272 !		— —		— —
Asian/Pacific Islander		327		310		307
Hispanic		282		283		282
White		312		309		310
Geometry						
Black		288		280		276
American Indian		284 !		— —		— —
Asian/Pacific Islander		328		319		314
Hispanic		292		288		288
White		314		312		312
Data analysis						
Black		278		278		276
American Indian		284 !		— —		— —
Asian/Pacific Islander		318		302		301
Hispanic		291		282		282
White		312		312		310
Algebra						
Black		288		281		279
American Indian		280 !		— —		— —
Asian/Pacific Islander		336		314 ‡		318
Hispanic		292		287		290
White		312		310		310

— — Indicates insufficient sample size to permit a reliable estimate.
‡ Indicates a significant difference between S1 and S2 results.
NOTE: ! Interpret with caution: the nature of the sample does not allow accurate determination of the variability of this statistic.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Chapter Summary

In the 1996 science assessment, the presence of accommodations for some students with disabilities and limited English proficient (LEP) students had little or no effect on the mean scale scores of any of the major reporting subgroups. There were, however, some differences in the mean scale scores for LEP students across the samples, a group for which results are currently not routinely reported. This may have important implications for reporting results for this subgroup in the future both within an assessment year and across assessment years.

The mathematics results at the fourth and eighth grades showed no significant differences between the three sample types (pre-1996 inclusion criteria, 1996 inclusion criteria, and 1996 criteria with accommodations). At the twelfth grade, there were only three significant differences. Two of these, between the S2 and S3 samples in geometry, overall and for female students, are not independent. The third significant difference is between S1 and S2 in algebra and functions for the Asian/Pacific Islander student group.

In summary it can be stated that the differences in the two sets of inclusion rules and the effects of including accommodated students were minimal on the mathematics and science assessments of 1996. As pointed out earlier, for the science assessment there were more significant differences (9) between groups in the two random half samples of the science assessment than between the samples with and without accommodations in both the science and the mathematics assessments. Although there were a few significant differences in average scores, none of them was associated with the main reporting groups for which data appear in the NAEP report cards, that is, the composite scores by the reporting groups described in detail earlier in this chapter. Furthermore, the differences observed in the mathematics assessment were not readily interpretable and make little sense when evaluated against the inclusion rate data presented in chapters 3 and 5.

The conclusion based on analysis of both the science and mathematics assessment data from the 1996 assessment year is that including students with accommodations and changing the rules for determining which students are included in the assessments had almost no measurable effect on the scale score results for main reporting groups. At two grades, however, science results for groups of LEP students with score distributions at the lower end of the scale were affected.

Chapter 8

Concluding Comments

As asserted in the introductory chapter, the inclusion of special needs students is a serious issue for educators and policymakers. Only with equitable representation of students with disabilities and students of limited English proficiency (LEP) can summary assessment data be truly representative of the entire nation. Inclusion of a representative proportion of special needs students in the NAEP assessments depends on the consistent application of locally implemented policy on who should participate and the provision of necessary accommodations and adaptations in testing for those who do participate.

Both the *NAEP 1996 Mathematics Report Card* and *the NAEP 1996 Science Report Card* contain a concluding chapter titled, "Exploring a More Inclusive NAEP." These chapters described the experiment, carried out within the traditional NAEP assessment, that involved revised inclusion criteria and the provision of accommodations and adaptations for special needs students. In addition to the core assessment activities that continued national trend reporting, the 1996 mathematics and science assessments included supplemental samples of schools and students. The supplemental samples were designed to allow the program to study the feasibility and impact of increasing the numbers of LEP students and students with disabilities who are included in NAEP. Revised inclusion rules were implemented in one sample, and assessment accommodations and adaptations were permitted in another. In all samples, questionnaires filled out by knowledgeable school staff were obtained for each sampled special needs student, regardless of whether the student was ultimately included in NAEP. The report card chapters presented some initial results on the impacts of the policy revisions on inclusion rates of special needs students, and promised a more detailed analysis in a future report. This report presents these additional analyses.

The NAEP 1996 report cards and each of the chapters in this report contain detailed summaries of results. A comprehensive summary will not be repeated here. However, it is useful to review the major findings on inclusion from the NAEP 1996 assessments. In addition to a general summary, this chapter will highlight a number of future directions for research on inclusion in NAEP.

General Summary

The findings from the 1996 research effort can be grouped into three major areas: 1) descriptive results on the background and educational experiences of special needs students; 2) the impacts of revised criteria and the provision of accommodations and adaptations on inclusion rates; and 3) the impacts of increased inclusion on the technical quality and characteristics of NAEP results.

Background and Educational Experiences of Special Needs Students

Chapters 2 and 4 presented descriptive statistics on background characteristics and educational experiences of the students with disabilities and the LEP students attending schools in NAEP's sampling frame.

Students with Disabilities. Students with disabilities who attended schools that fell within the NAEP sampling frame, a target population that did not include ungraded schools or special schools for the deaf and blind, were generally students with mild to moderate cognitive disabilities. "Learning disability" was by far the most frequently reported category for students with disabilities, with close to three of four students so identified at each of the three grades. About half of the students at each grade had mild disabilities. The remaining half at each grade were almost all categorized with moderate to severe disabilities. Very few students were categorized as profoundly disabled at all three grades; about half of all students with disabilities were mainstreamed in academic subjects at least 80 percent of the time.

A substantial proportion of students with disabilities had access to the general education curriculum. However, exposure to this general curriculum appeared to differ by grade and subject. Across the three grades, no more than half of the students were receiving grade-level instruction in reading/language arts. At grades 4 and 8, a larger percentage received grade-level instruction in mathematics and science than in reading/language arts. However, at grade 12 the percentages receiving grade-level instruction in mathematics and science more closely mirrored those in reading/language arts. In all three subject areas and at all three grades, almost all students receiving grade-level instruction received the same curriculum content as their nondisabled peers. However, such a situation was far less common among students with disabilities receiving below-grade-level instruction.

Judgments about the academic performance of students with disabilities are generally consistent with existing research in suggesting low levels of academic achievement for the majority of these students. In all three grades, more than 75 percent of students with disabilities were judged to be performing below grade level in reading/language arts. At grade 4, reported performance levels in mathematics and science were somewhat higher than those in reading/language arts. However, at the higher grades, a larger incidence of below-grade-level performance was reported. It should be noted, however, that the educational significance of results on instruction and performance levels could be better interpretered if comparable data for students without disabilities were available. Such information was not collected in NAEP.

The use of accommodations and adaptations in local testing of students with disabilities was fairly common. Across the three grades, respondents reported that 42 to 44 percent of students with disabilities received some form of accommodation or adaptation in testing. Between 33 and 41 percent of students with disabilities used presentation, timing, or setting accommodations. Response accommodations appeared to be used less frequently, ranging from 15 percent at grade four to 19 percent at grade twelve.

Limited English Proficient Students. LEP students who attend graded schools that fell within the NAEP sampling frame were a mixture of recent immigrants and long-term U.S. residents. The percentage who were recent immigrants tended to be greater at the higher grades. At grade 12, at least 56 percent of students with limited English proficiency had lived in this country for five years or less. No more than 44 percent of grade–four students have lived in this country for five years or less. At least 42 percent of fourth-grade LEP students had lived in the U.S. all their lives, while no more than 22 percent of twelfth–grade students have lived in the U.S. all their lives. Most LEP students spoke Spanish as their native language, with the next largest groups comprised of speakers of one of a number of Asian languages.

Most LEP students at all three grades had been enrolled for three or more years in a school where English is the primary language of instruction. However, the vast majority of LEP students at all three grades received some special instruction (87 percent of grade–four LEP students, 80 percent of grade–eight LEP students, and 81 percent of grade–twelve LEP students). Such special instruction was more common in reading/language arts than in mathematics and science. At grade 4, nearly one in four LEP students received academic instruction primarily in his or her native language. Among LEP students at the two higher grades, however, native language instruction in reading/language arts, mathematics, and science was far less common (8 percent or less). Revised NAEP inclusion criteria specify that all LEP students receiving academic instruction primarily in English for three or more years should be assessed in NAEP under standard conditions. At least 44 percent of grade–four LEP students, 47 percent of grade–eight students, and 65 percent of grade–twelve students met this criterion.

Questionnaire data suggest that, at all three grades, most LEP students taught in English experienced instruction that was at grade level in mathematics and science. At grade four, over 83 percent of LEP students were reported to be receiving instruction at their grade level in mathematics and in science. Close to three-quarters of LEP students at grade eight and 64 percent of LEP students at grade twelve received grade-level instruction in mathematics and science. At the lower two grades, English-language instruction appeared less common in reading/language arts. Despite these levels of instruction, at least two-thirds of grade–four and grade–eight LEP students receiving English-language instruction were judged as performing below grade level in reading/language arts; at grade twelve, at least 50 percent were so judged. In science, the percentages reported performing below grade level ranged from at least 30 percent at grade twelve to at least 47 percent at grades four and eight. In mathematics, the percentages performing below grade level ranged from at least 33 percent at grade twelve to at least 46 percent at grade eight.

Accommodations or adaptations are not routinely provided for LEP students. Respondents indicated that 37 percent of grade–four LEP students, 27 percent of grade–eight LEP students, and 22 percent of grade–twelve LEP students used accommodations and adaptations in achievement testing in their schools.

Impacts of NAEP Policy Changes on Inclusion Rates

The three-sample design used in 1996 allowed a separation of the effects on inclusion rates of 1) simply changing the wording of the NAEP policy on including/excluding students from the assessment (compared to the "old" policy), and 2) the new policy plus the provision of accommodations/adaptations during the assessment. Analyses were performed separately for students with disabilities and for limited English proficient students.

Students with Disabilities. Results presented originally in the NAEP 1996 report cards, and the follow-up results presented here indicate that an increase in inclusion rates was not achieved by simply changing the wording of the policy, as implemented in participating schools. Comparisons of the overall percentages of students with disabilities included in NAEP with the original and revised criteria provided no strong evidence of a difference. Further analysis among subgroups of students (e.g., by severity of disability, by percentage of time mainstreamed, and by grade-level of instruction) all corroborated the overall result. Moreover, the percentages of students with disabilities tested in NAEP equaled or exceeded the percentages of students judged by questionnaire respondents as capable of meaningful participation in NAEP without accommodations. These results suggest that further increases in the numbers of students with disabilities participating in NAEP are not likely to result solely from revisions to inclusion criteria.

In contrast, inclusion rates of students with disabilities were considerably increased at grades 4 and 8 when the inclusion policy change was combined with the availability of accommodations/adaptations during testing. A similar increase at grade 12, however, was not found. Generally, when accommodations and adaptations were made available, the observed rates of inclusion were in the low 70s at grades four and eight, and around 50 percent at grade twelve. Further analysis by subgroup at grade 4, and to a lesser extent at grade 8, revealed that the increases in inclusion rates were evident across a broad spectrum of students with disabilities. At grade 4, the provision of accommodations and adaptations increased inclusion among the following groups: students with mild and moderate disabilities; students mainstreamed less than 50 percent of the time; students mainstreamed more than 50 percent of the time; students receiving instruction at/or above grade level; and students receiving instruction below grade level. Though not always statistically significant, a similar pattern of broad-based impact was evident at grade 8.

The SD/LEP questionnaire, filled out by school personnel who knew the students, provided an opportunity to estimate the best possible inclusion rate, i.e., the percentage of special needs students who would be included if appropriate accommodations/adaptations were provided. These estimates ranged from a low of 66 percent at grade 12 to a high of 79 percent at grade 8. Comparisons of these questionnaire percentages to the actual inclusion percentage observed in the S3 sample suggest that students with disabilities in grade four are already being included at as high a rate as the questionnaire respondents judged reasonable, but that improvements may still be possible at grades 8 and 12. How to obtain such improvements is, however, far from clear. Further improvements in inclusion rates would likely involve: 1) a change in policy toward who would qualify for accommodations; and/or 2) an expansion in the types of special testing conditions provided.

LEP Students. Results presented originally in the NAEP 1996 report cards indicated little evidence that the revisions to the wording of the inclusion criteria, without the provision of accommodations and adaptations, had any impact on inclusion rates at grades 8 and 12. The evidence at grade 4 was less clear. Inclusion rates in S2 (the sample in which the revised criteria were used, but accommodations were not permitted) were 20 percent lower than in S1 (the sample in which the original criteria were used). This difference was not, however, statistically significant.

Additional follow-up analyses presented in this report do provide some evidence that the revised criteria, when used in the absence of accommodations and adaptations, would result in a decrease in inclusion rates at grade 4. For example, when LEP students were grouped by years enrolled in an English-language school (less than 2 years versus 2 years or more), inclusion rates were lower in S2 than in S1 for both groups of students. When LEP students were grouped by years receiving academic instruction primarily in English (less than 3 years versus 3 years or more), inclusion rates were lower for both groups of students. Among LEP students who would be best assessed in their native language, inclusion rates were dramatically lower in S2 than in S1. With one exception (the S2 sample at grade 8), the percentages of LEP students assessed in NAEP without accommodations and adaptations exceed estimates indicated by the questionnaire respondents. Thus, as was the case for students with disabilities, increases in inclusion are unlikely to result solely from revisions to the inclusion criteria.

When used in conjunction with the revised criteria, the provision of accommodations and adaptations did increase inclusion among LEP students in grade 4 and, to a lesser extent, grade 8 when compared to use of the revised criteria alone. Inclusion rates were higher in S3 than in S2 overall and for a wide range of subgroups. For example, the provision of accommodations and adaptations appeared to increase inclusion rates at grades 4 and 8 for LEP students regardless of the number of years spent in English-language schools. Grade–4 inclusion rates were higher in S3 than in S2 regardless of the number of years the student had been receiving academic instruction in English. Numerous other examples of increased inclusion rates are evident in the tables presented in chapter 5.

However, the evidence is less compelling as to whether the combined effect of the two changes resulted in increased inclusion rates. Overall inclusion rates appear higher in S3 than in S1 at grades 4 and 8. These apparent differences are not, however, statistically significant. Grade 4– and grade–8 analyses by subgroup reveal a number of cases where S3 rates appeared higher than S1 rates, but a very small number of these were statistically significant. Moreover, the subgroup analyses also revealed instances where there were no apparent differences.

As was the case for students with disabilities, the SD/LEP questionnaire, filled out by school personnel who knew the students, provided an opportunity to estimate the best possible inclusion rate. In other words, the questionnaire responses provided the opportunity to estimate the percentage of LEP students who would be included if appropriate accommodations and adaptations were provided. These estimates ranged from a low of 89 percent at grade four to a high of 96 percent at grade eight. Comparisons of these questionnaire percentages to the actual inclusion percentages suggest that further improvements in inclusion might still be possible if the list of permitted accommodations and adaptations could be expanded. Since the most effective addition would likely involve native language testing in languages other than Spanish, the cost of such an approach remains a serious obstacle.

Impacts on Technical Quality

In addition to the inclusion question as discussed in terms of percentage changes in student inclusion rates, psychometric questions about the possible effects on the NAEP scale of including greater numbers of special needs students with accommodations and adaptations in testing were addressed in chapters 6 and 7.

The technical quality issues addressed in this report were as follows.

- The Item Response Theory (IRT) methods used by NAEP assume that a common IRT model fits the data for all the major subgroups reported in NAEP. This raises an important question about whether the data obtained from special needs students tested with accommodations and adaptations can be fit with the same IRT models as data obtained from students tested under standard conditions.
- A related question concerns the aggregate effect of such data on NAEP IRT scaling. Does the presence of a small percentage of misfitting data have any measurable impact on the technical characteristics (item-level model fit, test characteristic curves, and test information curves) of aggregate NAEP scaling results?
- A third question concerns the impact of including data from students tested with accommodations and adaptations on the group-level proficiency statistics reported in NAEP. Specifically, do group-level statistics that include data from such nonstandard administrations differ in any meaningful way from the statistics that would be reported if only standard administration data were included?

Can Results from Nonstandard Administration be Fit With the Overall NAEP IRT Model?

DIF analyses and judgments of item-level model fit provide some evidence that data from nonstandard administrations cannot be fit with the same IRT model as data from standard administrations. Such evidence was clearer for the science assessment than for the mathematics assessment. Despite this item-level evidence, there was little to suggest that more global aspects of NAEP scaling results (item-level model fit indicators, test characteristic curves, or test information curves) or the group-level statistics reported in NAEP (i.e., average scale score results for typical NAEP reporting groups) were affected in any meaningful way.

Differential Item Functioning Analysis. DIF, normally used to check on unusual differences in item performance, conditioned on total score performance, among minority versus majority (reference) groups was employed here as one method to examine the effect of accommodated performance versus regularly assessed performance.

At two of the three grades in science, there is some evidence of accommodations-related DIF. Such evidence is noteworthy, given the small sample sizes involved. In contrast, the mathematics assessment provided little evidence of accommodations-related DIF.

Judgments of Item-Level Model Fit. Judgments of model fit were made based on a graphical examination of the fit of the accommodation items to the IRT model for two groups of students; 1) those who were administered the test with accommodations; and 2) those assessed with standard conditions. Empirical item characteristic curves (ICCs) were compared to the operational or theoretical ICCs used in reporting the NAEP 1996 results. Comparisons were conducted for students tested with accommodations and for those tested under standard conditions. For both mathematics and science, there was evidence that the match between empirical and theoretical ICCs was worse for the accommodated student data. In science, this finding is consistent with DIF results in suggesting that a common model may not fit data from both groups. In mathematics, these graphical results are somewhat inconsistent with the DIF results where few differences were observed.

Did the Revision of Inclusion Criteria and Provision of Accommodations Have a Measurable Impact on Aggregate NAEP Scaling Results?

Distributions of Item-Level Model Fit Indices. A standard by-product of NAEP's operational scaling procedure is the production of item-level model fit indices. These indices measure how closely student responses conform to the model-based item response functions estimated for each item. Separate IRT scalings were carried out for the S1, S2 and S3 samples, and distributions of the indices across the item pool were compared by the different sample types. To provide a reference point, the science S2 sample was divided into two half samples. Separate scalings were conducted in the half sample and the results were comparable. Fit statistic results for the items were categorized by size into five mutually exclusive categories, and chi-square testing to compare distributions was done under the assumption of equal category proportions. Only for grade–8 science were any differences found. There they had actually appeared better with the inclusion of accommodated students in S3. In mathematics no significant differences were found.

Test Characteristic and Test Information Curves. The test characteristic functions show expected levels of performance conditional on subject matter proficiency, and indicate how well the measurement instrument differentiates between students having slightly different values on the scale. The test characteristic curves from the separate IRT scalings were plotted and compared for the different samples. ***Overall, the differences between test characteristic and test information curves in science based on the S2 and S3 sample are small and look no greater than those obtained in random half samples of S2. Hence, there is little evidence that the presence of accommodated students has changed the nature of the scales.*** There was more variability among the mathematics subscale curves, because of the smaller numbers of items contained in the scales. However, the curves are quite similar across the samples representing different inclusion criteria and accommodations at all three grade levels.

Did Inclusion of Results from Nonstandard Administration Affect Overall Proficiency Statistics?

Proficiency Means. NAEP traditionally reports proficiency means by a constant set of reporting groups: gender; race/ethnicity; parents' education; type of school; region of the country; type of location; Title 1 participation; and, eligibility for the National School Lunch Program. The appropriate mean proficiency scores for each reporting variable were compared for the S1, S2, and S3 samples, by grade level for both the science and mathematics assessments. Similar comparisons were made between the random half samples in science. The few significant differences that were found were: 1) not associated with the main reporting groups for which data appear in the NAEP report cards; or 2) not readily interpretable given the inclusion rates results reported in earlier chapters. In fact, for the science assessment, there were more significant differences (9) in group-level results between groups in the two random half samples than between the samples with and without accommodations in either mathematics or science (5). There were, however, some differences in the mean scale scores for LEP students across the samples. This may have important implications for reporting results for this subgroup both within an assessment year and across assessment years.

In total there was little evidence that changing inclusion rules or including accommodated students' data in the scaling had a measurable impact on the scale score results for major reporting groups in either science or mathematics. There was, however, some indication from the science analyses that results for groups consisting primarily of LEP students were affected by the change in criteria and the provision of accommodations.

Directions for Future NAEP Research

Results reported in chapters 3 and 5 suggested that, absent the provision of accommodations and adaptations, the change in inclusion criteria introduced experimentally in the 1996 mathematics assessment had, at most, a modest effect on inclusion percentages for students with disabilities and LEP students. Moreover, other results from these chapters indicate that, absent accommodations and adaptations, NAEP is already including as many special needs students under either the old or revised criteria as can be meaningfully tested. As a result of these findings, the program has already switched over to using the revised (i.e., S2 criteria) for its official reporting samples beginning in 1998.

The analyses on the technical characteristics of scale score results presented in chapters 6 and 7 also suggest that the additional procedural change of permitting students to test with accommodations and adaptions would not significantly affect the NAEP scale score results. If this is indeed the case, it may be possible for the program to achieve its joint goals of increasing inclusion of special needs students to the fullest degree possible while maintaining meaningful trend lines to past assessments, where more restrictive inclusion policies were in place. However, as with any research endeavor, it is important to acknowledge the limitations of the current study and to point to some profitable directions for future research.

One obvious limitation of the current study is that the 1996 assessment involved only two specific subject matter areas, mathematics and science. The descriptive results presented in chapters 2 and 4 suggest that instructional practices and estimated levels of performance for special needs students are different in the areas of reading/language arts than they are in mathematics and science. Students with disabilities are more likely to be receiving grade-level instruction in mathematics and science than in reading/language arts. LEP students are more likely to be receiving specially designed instruction in reading/language arts than in mathematics or science. Moreover, the list of reasonable accommodations and adaptations is likely subject specific. The mathematics assessment included a Spanish/English bilingual test booklet. Offering such a booklet in an assessment of reading or writing in English may not be defensible. It does not seem prudent to assume that the findings on the impacts on inclusion rates and the psychometric characteristics of the NAEP assessment instrument reported here necessarily generalize to all subjects, in particular to reading and writing assessments.

An additional limitation on the generality of the result reported here is that state policies and practices regarding the inclusion of special needs students are themselves evolving. As noted earlier in the report, the 1997 amendments to IDEA require states to include students with disabilites in statewide testing, offering accommodations or alternate testing situations as necessary. As states move their own testing programs and practices into compliance with these amendments, the number of students with disabilities who are included in regular statewide testing and are receiving some form of accommodation may increase. Such increases will no doubt "spill over" to the NAEP program and raise questions about the generality of the results reported here in a variety of ways. For example, the analyses in chapter 6 indicated that overall scaling results were not affected by the inclusion of data from students receiving accommodations. However, an effect might emerge in the future if the amount of such data were to increase. Similarly, the absence of differences in overall and subgroup scale score results across the S1, S2, and S3 samples reported in chapter 7 may not hold up under increases in the numbers of students tested with accommodations and adaptations.

The NAEP program has already had some experience with how changes in state assessment policies can impact NAEP results. On March 4, 1999, NCES released the public school results of the 1998 state NAEP assessment in reading. State NAEP assessments were conducted in 1998 at grade 4 in 44 jurisdictions and at grade 8 in 41 jurisdictions. The grade 8 state NAEP assessments in reading were the first of their kind. However, state NAEP assessments at grade 4 had also been conducted in 1992 and 1994. Four of the jurisdictions that participated in both the 1994 and 1998 NAEP reading assessments showed substantial increases in the percentages of students with disabilities who were excluded from the NAEP assessment.[1] The official NAEP reading results in 1998 were based on the S2 inclusion criteria, which did *not* allow for accommodations or adaptations. Indeed, the S2 inclusion criteria in NAEP specifically state that one acceptable reason for exclusion is that the student is normally tested with an accommodation or adaptation. Thus, one possible reason for the increase in exclusion seen in certain states in the most recent NAEP reading assessments is a disjuncture between assessment policies in NAEP and in state assessments regarding the use of accommodations. When NAEP switches over to reporting samples that allow accommodations, it is reasonable to expect that the numbers of students with disabilities testing under nonstandard conditions will increase.

As discussed in chapter 6, DIF and model fit analyses were conducted to examine the degree to which a common IRT model fits the data from both standard testing sessions and from those sessions in which accommodations were permitted. Because of sample size limitations, data from students with disabilities and limited English proficient students were combined, and data from accommodation sessions were pooled across types of accommodations. A more satisfactory analysis would avoid combining such disparate groups. As the data in chapters 2 and 4 show, students with disabilities and limited English proficient students differ dramatically in terms of their background characteristics and educational experiences. Analyses and model fit studies involving groups more homogenous with respect to these background factors are more likely to show interesting patterns of differential item functioning.

In addition, the DIF and model fit analyses compared results from all standard testing sessions to those from sessions involving accommodations. The former group is made up primarily of data from students who are neither students with disabilities nor LEP students and who typically perform at higher levels than special needs students. As the data in chapter 2 indicate, many students with disabilities are receiving below grade-level instruction, as well as curriculum content that differs from their nondisabled peers. As a result, the small amounts of DIF and lack of model fit detected here could be the result of curriculum effects, the impact of accommodations, general lack of model fit among low-performing groups, or some combination of these factors. A more satisfactory set of DIF analyses would attempt to establish a reference group for accommodations-related DIF analyses that is similar in performance and instruction levels to students receiving accommodations. However, such analyses would require much larger sample sizes than were available from the 1996 NAEP assessment

Moreover, in a study of the Kentucky Instructional Results Information System (KIRIS), Koretz[2] found results that suggest that the quality of scores obtained with assessment accommodations may depend on the particular accommodation in question. For example, the average score of fourth-grade, learning-disabled students with certain combinations of accommodations were well above those of students without disabilities. Similarly, the average scores of fourth-grade students with mild mental retardation given certain accommodations were only 0.1 standard deviations below the average for nondisabled students in reading and

[1] J. Mazzeo, J. Donoghue, & C. Hombo. (Memoranum, May 12, 1999). A summary of initial analyses of 1998 state NAEP inclusion rates.

[2] Koretz, D. (1997). *The assessment of students with disabilities in Kentucky.* (CSE Technical Report No. 431). Los Angeles, CA: Institute on Education and Training, CRESST/RAND.

0.1 standard deviations above the average in science—possibly unreasonable results, given that students with mental retardation by definition have generalized cognitive deficits. Of the accommodations recorded in KIRIS, providing students with the opportunity to dictate responses (offered to more than half of the learning-disabled fourth graders assessed) had by far the strongest positive association with scores.[3]

More recently, Trimble[4] reported that student performance on KIRIS with accommodations is generally lower than the performance of general students. In only a small number of instances (4 of 104 accommations per subject matter groupings) were the average scores of students using accommodations higher than the performance of the total group of students. As with the Koretz report, accommodations involving paraphrasing and dictation were among this small number of instances, perhaps suggesting the need for further examination of these particular accommodations. However, both these reports of Kentucky results results argue strongly for the need in the future for separate DIF and model fit analyses by type of accommodation.

Fortunately for NAEP, the 1998 assessment in reading will afford an opportunity to examine the generality of the results obtained in 1996, as well as provide the possibility of conducting some of the psychometric analyses reported here separately for students with disabilities and LEP students and, possibly, within accommodation category. The 1998 reading assessments again included a multisample design that will permit an investigation of the impacts of providing accommodations on inclusion rates in reading. In addition, state assessments conducted at grades 4 and 8 used a similar, split-sample design in which the accommodations and adaptations were permitted in half of the schools within each participating jurisdiction. State-by-state analyses will allow an examination of state-level impacts of providing accommodation on inclusion rates. In addition, the aggregate samples across all participating jurisdictions may provide sufficient sample sizes to allow for psychometric analyses in more meaningful and homogenous subgroups.

Several analyses, reported on in chapters 3 and 5, raise some questions about the degree to which NAEP's intended inclusion policies are actually implemented. Using both original and revised criteria, students with disabilities were to be excluded from NAEP on the basis of their IEP, or on the basis of school judgments that the student was too severely cognitively impaired to be meaningfully tested. Analyses presented in chapter 3 indicate that most inclusion decisions were made on the basis of what is stated in the IEP, and relatively few exclusion decisions were made on the basis of a judgment of severe cognitive impairment, absent corroborating direction from the IEP. However, the results also suggest that, for substantial percentages of excluded students, neither determination by the IEP team nor the presence of cognitive impairments was given as reason for exclusion. The absence of a normally used accommodation may account for some percentage of these exclusions. However, other factors may be at work. Additional follow-up research into what these other reasons for exclusion were could benefit the program in crafting future changes to policies and procedures.

---

[3] Ibid., p.vi.

[4] Trimble, S. (1998). *Performance trends and use of accommodations on a statewide assessment: Students with disabilities in the KIRIS on-demand assessments from 1992-93 through 1995-96*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

The revised criteria for LEP students stated that students receiving three or more years of academic instruction primarily in English should be included in NAEP. In general, inclusion rates were higher among students with three or more years of academic instruction in English than for students with less than three years. This pattern of results was evident, with one or two exceptions, over all grades and samples. However, it was also evident that not all students reported to be receiving academic instruction in English for more than three years were included in the assessment. The percentages of such students included in NAEP ranged from 92 (at grade twelve in the S1 and S3 samples) to 67 (at grade four, in the S2 sample). These less-than-perfect levels of inclusion could be the result of a number of factors, including incorrect application of the inclusion criteria and invalid or incorrect questionnaire responses. Additional field research into the reasons for these apparent discrepancies would clearly benefit the program in improving questionnaire design and ensuring more consistent application in the field procedures.

NAEP's accommodations policy is built around the concept of testing students under the conditions that they typically encounter in their classroom, school, district, or state-level testing. As part of the SD/LEP questionnaire, respondents were asked to indicate for each student whether accommodations or adaptations are used for achievement testing. The results for this question for students with disabilities and LEP students revealed inconsistencies that point to important methodological issues. In the S3 samples, substantial percentages of students with disabilities and LEP students who normally receive accommodations or adaptations were tested under standard conditions. While some of these students may have used accommodations or adaptations not readily provided by the program, it is unlikely that this is the case for all such students. Furthermore, a relatively small percentage of students who do not regularly receive accommodations or adaptations used them in NAEP when they were made available. As noted earlier, some students who do not normally test with accommodations may be given this opportunity because, in the judgment of school staff, they can more meaningfully demonstrate what they know and can do. Nonetheless, such a practice does not comply with the current NAEP program policy on who is eligible to receive accommodations. Moreover, these inconsistencies could also at least partly be the result of invalid data provided on the SD/LEP questionnaire, or be indicative of confusion and improper implementation of field procedures. Follow-up research in future assessments will be required to resolve the reasons for such inconsistencies.

The National Assessment of Educational Progress will continue to explore ways to increase the participation of special needs students in the assessment, test those students in ways that allow them to fairly demonstrate what they know and can do, and reflect the performance of such students in the results it reports. The program is committed to accomplishing these changes while maintaining to the greatest degree possible its ability to report performance trends over time. An initial step toward these goals was taken with the design of the1996 assessment and subsequent assessments have been designed to permit continued progress in these areas. Much was learned from the the experiment conducted in the 1996 assessment, but much more needs to be, and will be, learned from the research activities planned for 1998 and beyond.

Appendix A

Analysis of Integrity of the Special Needs Students Questionnaire Data

During the NAEP 1996 assessments in mathematics and science, SD/LEP questionnaires were distributed at each participating school. A copy of the full questionnaire is provided in appendix B. The SD/LEP questionnaire contained two major sections: (1) a 30-question section to be completed for all students in the sample who were eligible for special education services under the Individuals with Disabilities Education Act (IDEA), or as a result of section 504 of the Rehabilitation Act, and, (2) a 36-question section to be completed for all students in the sample who were identified by their schools as LEP students. A knowledgeable school staff member (e.g., a member of the student's IEP team in the case of students with disabilities or a bilingual education or ESL teacher in case of LEP students) was asked to fill out the SD/LEP questionnaire for each student with a disability or for each LEP student, regardless of whether that student was judged capable of participating in the NAEP assessment.

Questionnaires were returned for the large majority of students with disabilities and LEP students. Tables A1.a and A1.b present sample sizes, numbers of students matched to questionnaires, and percentages of matched and missing data for students with disabilities and LEP students, respectively. As shown in table A1.a, questionnaire match rates for students with disabilities in the mathematics samples ranged from 82 percent at grade 4 to 78 percent at grade 12. Match rates for LEP students in table A1.b ranged from 76 percent at grade 4 to 68 percent at grade 12. Such rates are below NCES statistical standards for survey response rates, somewhat lower than those typically observed for the teacher questionnaires that are routinely administered in NAEP, and substantially lower than those typically observed for the NAEP school and student questionnaires. These lower-than-expected match rates raise potential concerns about the representativeness of the results presented in chapters 2 through 5.

Table A1.a — Students with disabilities questionnaire outcomes — N's, percentages matched, and percentages of missing data: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD

	Number of students with disabilities			
	Total	Matched to a questionnaire	Percentage matched to a questionnaire	Percentage missing questionnaire
Grade 4				
Mathematics	1,194	978	82	18
Science	1,269	1,112	88	12
Grade 8				
Mathematics	1,391	1,114	80	19
Science	1,367	1,098	80	19
Grade 12				
Mathematics	1,008	784	78	22
Science	917	740	81	19

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table A1.b — Students with limited English proficiency questionnaire outcomes — N's, percentages matched, and percentages of missing data: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD

	Number of students with limited English proficiency			
	Total	Matched to a questionnaire	Percentage matched to a questionnaire	Percentage missing questionnaire
Grade 4				
Mathematics	754	570	76	24
Science	1,024	790	77	23
Grade 8				
Mathematics	650	456	70	30
Science	656	482	73	27
Grade 12				
Mathematics	508	347	68	32
Science	538	373	69	31

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

In order to examine the degree to which the available data were representative of the full NAEP population of students with disabilities and LEP students, analyses were conducted that compared the subset of students with returned questionnaires to all such students with respect to selected school and student demographic characteristics. Separate analyses were conducted for students with disabilities and for LEP students. The analyses were carried out without the use of sampling weights. Unweighted analyses are appropriate in this context since the question being investigated is: Conditional on being selected into the sample, can the missing questionnaires be assumed to be a random subset of the entire group of special needs students sampled for the assessment?

The results of both sets of analyses are presented in this appendix. The general findings are that the sample of students matched to questionnaires differed little from the full set of students on all the variables examined. The findings were similar for students with disabilities and for LEP students. A more detailed discussion appears below. Based on the results presented here, there is little evidence to suggest that the data from the subset of students matched to questionnaires is not representative of the full sample.

School-level Variables

In order to examine whether the returned questionnaires came from a nonrepresentative subset of schools, three different school-level variables were examined:

1. **School average NAEP mathematics scores** – the average NAEP mathematics score for each school in the sample. The averages were derived from number-correct book-level scores, which were transformed to have a normal distribution with a mean of zero and a standard deviation of 1 in the full NAEP 1996 mathematics sample.
2. **Percent Black students in school** – the percentage of black students enrolled in the school, as given in the *Quality Education Data* (QED) [1] data file.
3. **Percent Hispanic students in school** – the percentage of Hispanic students enrolled in the school, as given in the QED data file.
4. **School Location** – a variable provided to ETS by the sampling contractor, Westat.

Tables A2.a and A2.b show the average of the first three of these school-level variables for (1) all students with disabilities; and, for (2) students with disabilities matched to a SD/LEP questionnaire. At all three grades, students with disabilities who were matched to questionnaires came from schools that, on average, performed slightly better on the NAEP mathematics test than did the full sample. The same was true of LEP students at grades 4 and 8. At grade 12, LEP students matched to a questionnaire were from schools that performed slightly worse on the NAEP test than did the full sample. For students with disabilities and LEP students at all three grades, there was almost no difference between the matched and full sample with respect to the percent of minority enrollment in the schools.

[1] The Quality Education Data file is a commercially available listing of school- and district-level data for schools in the United States. Data from this file on the schools sampled for the 1996 NAEP were provided to ETS by its subcontractor, Westat.

Table A2.a — Students with disabilities questionnaire outcome – percentages of total versus matched to a questionnaire for mathematics performance and ethnicity by grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

	Mathematics	
	Percentage of all students with disabilities	Percentage of students with disabilties matched to questionnaire
Grade 4		
Average school mathematics score	– .14	–.11
Percent Black students in school	21	22
Percent Hispanic students in school	12	11
Grade 8		
Average school mathematics score	–.11	–.09
Percent Black students in school	21	22
Percent Hispanic students in school	12	10
Grade 12		
Average school mathematics score	–.21	–.17
Percent Black students in school	21	20
Percent Hispanic students in school	10	10

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table A2.b — Limited English proficient students questionnaire outcomes – percentages of total versus matched to questionnaire for mathematics performance and ethnicity by grade: 1996 NAEP mathematics sample

	Mathematics	
	Percentage of all LEP students	Percentage of LEP students matched to a questionnaire
Grade 4		
Average school mathematics score	– .59	–.56
Percent Black students in school	13	13
Percent Hispanic students in school	50	49
Grade 8		
Average school mathematics score	–.48	–.43
Percent Black students in school	14	12
Percent Hispanic students in school	41	41
Grade 12		
Average school mathematics score	–.54	–.56
Percent Black students in school	16	18
Percent Hispanic students in school	39	38

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Tables A3.a and A3.b show similar results for the school location variable. For students with disabilities, there were slight differences between the full and matched samples at all three grades. The percentage of students from rural schools was slightly higher in the matched sample than in the full sample at all three grades. The percentage of students from urban schools was slightly lower in the matched sample than in the full sample at grades 8 and 12. For LEP students, the percentage distributions by school location did not differ at grade 8. At grade 4, the percentage of students from urban schools was slightly lower and the percentage of students from suburban schools was slightly higher in the matched sample than in the full sample. At grade 12, the percentage of students from urban schools was slightly higher and the percentage of students from suburban schools was slightly lower in the matched sample than in the full sample.

Table A3.a — Students with disabilities questionnaire outcomes – percentages of total versus matched to questionnaire by type of school location and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

	Mathematics	
School location	**Percentage of all students with disabilities**	**Percentage of students with disabilties matched to questionnaire**
Grade 4		
Urban	38	38
Suburban	38	37
Rural	23	25
Grade 8		
Urban	38	36
Suburban	38	38
Rural	25	26
Grade 12		
Urban	36	33
Suburban	33	33
Rural	31	34

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table A3.b — Limited English proficient students questionnaire outcomes – percentages of total versus matched to questionnaire by type of school location and grade: 1996 NAEP mathematics sample

School location	Mathematics: Percentage of all LEP students	Mathematics: Percentage of LEP students matched to a questionnaire
Grade 4		
Urban	58	56
Suburban	37	40
Rural	5	4
Grade 8		
Urban	61	61
Suburban	34	33
Rural	5	6
Grade 12		
Urban	63	67
Suburban	33	29
Rural	5	4

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Student-Level Variables

In order to examine whether the returned questionnaires came from non-representative subsets of students with disabilities and LEP students, four student-level variables provided for each student by Westat field staff were examined: (1) gender, (2) race/ethnicity, (3) whether the student was receiving Title I services, (4) eligibility for free or reduced-price lunch. The percentages of assessed and excluded students for each matched and full sample were also examined.

Tables A4.a and A4.b present the gender results. There were no differences between matched and full samples in the percentages of males and females for students with disabilities or LEP students.

Table A4.a — Students with disabilities questionnaire outcomes – percentages of total versus matched to a questionnaire by gender and grade: 1996 NAEP mathematics sample

Gender	Percentage of all students with disabilities	Percentage of students with disabilties matched to a questionnaire
Grade 4		
Male	65	65
Female	35	35
Grade 8		
Male	65	64
Female	35	36
Grade 12		
Male	67	66
Female	33	34

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table A4.b — Limited English proficient students questionnaire outcomes – percentages of total versus matched to a questionnaire by gender and grade: 1996 NAEP mathematics sample

Gender	Percentage of all LEP students	Percentage of LEP students matched to a questionnaire
Grade 4		
Male	51	51
Female	49	49
Grade 8		
Male	57	54
Female	43	46
Grade 12		
Male	54	52
Female	46	48

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table A5.a and A5.b present the race/ethnicity results. For students with disabilities at all three grades, the percentage of white students was slightly higher in the matched sample than in the full sample. Conversely, the percentage of Hispanic students was slightly lower in the matched sample than in the full sample. For LEP students, the percentage distributions were nearly identical for the matched and full sample with one exception. At grade 4, the percentage of Asian/Pacific Islander students was slightly higher in the matched sample than in the full sample.

Table A5.a — Students with disabilities questionnaire outcomes – percentages of total versus matched to a questionnaire by race/ethnicity and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD

Race/ethnicity	Percentage of all students with disabilities	Percentage of students with disabilties matched to a questionnaire
Grade 4		
White	59	62
Black	23	24
Hispanic	14	11
Asian/Pacific Islander	2	1
Native American	2	1
Other	0	0
Grade 8		
White	59	61
Black	25	25
Hispanic	13	11
Asian/Pacific Islander	2	1
Native American	1	1
Other	0	1
Grade 12		
White	57	60
Black	28	27
Hispanic	12	11
Asian/Pacific Islander	1	1
Native American	1	2
Other	0	0

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table A5.b — Limited English proficient students questionnaire outcomes – percentages of total versus matched to questionnaire by race/ethnicity and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

Race/ethnicity	Percentage of all LEP students	Percentage of LEP students matched to a questionnaire
Grade 4		
White	5	4
Black	4	3
Hispanic	72	73
Asian/Pacific Islander	16	19
Native American	2	2
Other	0	0
Grade 8		
White	13	13
Black	4	2
Hispanic	65	66
Asian/Pacific Islander	18	19
Native American	0	0
Other	0	0
Grade 12		
White	11	10
Black	10	10
Hispanic	50	51
Asian/Pacific Islander	26	25
Native American	2	3
Other	1	1

SOUCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Tables A6.a and A6.b present results on the percentage of students receiving Title I services. NAEP collects this information from school records for all sample students. For students with disabilities and LEP students at grades 4 and 8, there was almost no difference between the full and matched samples in the percentage of students receiving these services. There were, however, small differences at grade 12 for both groups. For students with disabilities, the percentage of students receiving Title I services was slightly lower in the matched sample than in the full sample. For LEP students at grade 12, the percentage of students receiving Title I services was slightly higher.

Table A6.a — Students with disabilities questionnaire outcomes – percentages of total versus matched to a questionnaire by whether student is receiving services by Title I and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

Receiving Title I Services	Percentage of all students with disabilities	Percentage of students with disabilties matched to a questionnaire
Grade 4		
Yes	32	33
No	68	67
Grade 8		
Yes	15	14
No	85	86
Grade 12		
Yes	6	4
No	94	96

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

THE NATION'S REPORT CARD naep

Table A6.b — Limited English proficient students questionnaire outcomes – percentages of total versus matched to a questionnaire by whether student is receiving services by Title I and grade: 1996 NAEP mathematics sample

Receiving Title I Services	Percentage of all LEP students	Percentage of LEP students matched to a questionnaire
Grade 4		
Yes	63	64
No	37	36
Grade 8		
Yes	43	41
No	57	59
Grade 12		
Yes	27	30
No	73	70

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Tables A7.a and A7.b present results on the percentages of students eligible to receive free or reduced-price lunch. NAEP collects this eligibility information from the schools for all sampled students. For students with disabilities and LEP students at grade 4 there was almost no difference between the full and matched samples in the percentage of students eligible. There were, however, small differences at grades 8 and 12. For students with disabilities at grades 8 and 12, the percentage of students not eligible was higher, and the percentage of students with information not available was lower in the matched sample than in the full sample. For LEP students at grades 4 and 8 there was almost no difference between the full and matched samples in the percentages of students eligible. For LEP students at grade 12, the percentage of students eligible was higher and the percentage with information not available was lower in the matched sample than in the full sample.

Table A7.a — Students with disabilities questionnaire outcome – percentages of total versus matched to a questionnaire by free/reduced price lunch eligibility and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD NAEP

Eligibility for free/reduced lunch	Percentage of all students with disabilities	Percentage of students with disabilties matched to a questionnaire
Grade 4		
Eligible	51	51
Not eligible	37	38
Information not available	12	12
Grade 8		
Eligible	43	42
Not eligible	41	44
Information not available	16	14
Grade 12		
Eligible	29	30
Not eligible	53	58
Information not available	19	12

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table A7.b — Limited English proficient students questionnaire outcomes – percentages of total versus matched to a questionnaire by free/reduced price lunch eligibility and grade: 1996 NAEP mathematics sample

Eligibility for free/reduced lunch	Percentage of all LEP students	Percentage of LEP students matched to a questionnaire
Grade 4		
Eligible	77	78
Not eligible	14	12
Information not available	10	10
Grade 8		
Eligible	64	64
Not eligible	12	11
Information not available	24	25
Grade 12		
Eligible	42	46
Not eligible	21	23
Information not available	37	31

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Tables A8.a and A8.b present results on the percentage of students assessed and excluded. For students with disabilities and LEP students at grade 4, slightly fewer students were assessed in the matched sample than in the full sample. At grade 8, the reverse was the case for students with disabilities. For grade 12 students with disabilities and LEP students at grades 8 and 12, the percentages assessed were nearly identical in the matched and full samples.

Table A8.a — Students with disabilities questionnaire outcomes – percentages of total versus matched to a questionnaire by inclusion status and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

Inclusion Status	Percentage of all students with disabilities	Percentage of students with disabilties matched to a questionnaire
Grade 4		
Excluded	40	43
Assessed	60	57
Grade 8		
Excluded	41	38
Assessed	59	62
Grade 12		
Excluded	55	56
Assessed	45	44

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table A8.b — Limited English proficient students questionnaire outcomes – percentages of total versus matched to a questionnaire by inclusion status and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD NAEP

Inclusion status	Percentage of all LEP students	Percentage of LEP students matched to a questionnaire
Grade 4		
Excluded	44	48
Assessed	56	52
Grade 8		
Excluded	35	35
Assessed	65	65
Grade 12		
Excluded	25	26
Assessed	75	74

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Appendix B

SD/LEP Student Questionnaire

SD/LEP STUDENT QUESTIONNAIRE

POSITION OF PERSON
COMPLETING QUESTIONNAIRE

- Ⓐ Principal/Assistant Principal
- Ⓑ Special Education Teacher
- Ⓒ Bilingual Education/ESL Teacher
- Ⓓ Classroom Teacher
- Ⓔ Other (specify) ________________

A representative sample of students across the country, including some students in your school, have been selected to take part in the National Assessment of Educational Progress (NAEP). The current assessment focuses on mathematics and science. As part of the assessment, NAEP will investigate the relationship between students' achievement and various school, teacher, and home factors that may influence this achievement. In order to obtain a complete picture of how **all** children are doing, it is important to collect information on **all** those students who have been identified as having a disability, or limited English proficiency whether they will be assessed or NOT.[1] We are asking you to complete this questionnaire about one of those students.

We realize you are very busy; however, we urge you to complete this questionnaire as carefully as possible. The information you provide will be kept confidential.

NAEP is authorized under Public Law 103-382. While your participation is voluntary, your responses to these questions are needed to make this survey accurate and complete.

Please answer directly on the questionnaire with a number 2 pencil by gridding the appropriate letter or by writing your response in the space provided. Please complete questions 1 and 2 first. When you are finished, please return the questionnaire to your school's NAEP coordinator.

Thank you very much for your help.

1. Does this student have a disability (physical and/or mental)?

- Ⓐ No
- Ⓑ Yes (Please complete SECTION A.)

2. Does this student have limited English proficiency (LEP)?

- Ⓐ No
- Ⓑ Yes (Please complete SECTION B.)

If the student has both a disability and limited English proficiency, please complete SECTIONS A and B.

[1]For the purposes of this questionnaire, students with a disability include those who have an IEP or equivalent classification, such as those identified as part of the 504 program.

SECTION A: STUDENTS WITH DISABILITIES

Complete this section for all students with a disability who have an IEP or an equivalent classification.

3. Does this student have an Individualized Education Plan (IEP) or equivalent classification?

   Ⓐ Yes, IEP

   Ⓑ Yes, equivalent classification (define)

3a. Has the IEP team or an equivalent group determined that the student cannot participate in an assessment such as NAEP?

   Ⓐ No

   Ⓑ Yes

3b. Is this student's cognitive functioning so severely impaired that he/she cannot participate in this assessment?

   Ⓐ No

   Ⓑ Yes

4. Which of the following describes this student's disability? **(Grid in all that apply.)**

   Ⓐ Mental or cognitive impairment

   Ⓑ Hard of hearing

   Ⓒ Deafness

   Ⓓ Speech/language impairment

   Ⓔ Visual impairment/blindness

   Ⓕ Emotional disturbance

   Ⓖ Orthopedic impairment

   Ⓗ Learning disability

   Ⓘ Autism

   Ⓙ Traumatic brain injury

   Ⓚ Other (specify) ________________

5. What is the degree of this student's disability?

- Ⓐ Profound
- Ⓑ Severe
- Ⓒ Moderate
- Ⓓ Mild

6. What percentage of time is this student mainstreamed (i.e., with his/her nondisabled peers) in academic subjects (e.g., mathematics, reading/language arts, science)?

- Ⓐ 0%
- Ⓑ 10%
- Ⓒ 20%
- Ⓓ 30%
- Ⓔ 40%
- Ⓕ 50%
- Ⓖ 60%
- Ⓗ 70%
- Ⓘ 80%
- Ⓙ 90%
- Ⓚ 100%

7. For what percentage of time in the total school day is this student served by a special education program (both in a regular class with his/her nondisabled peers and outside the regular class)?

- Ⓐ 0%
- Ⓑ 10%
- Ⓒ 20%
- Ⓓ 30%
- Ⓔ 40%
- Ⓕ 50%
- Ⓖ 60%
- Ⓗ 70%
- Ⓘ 80%
- Ⓙ 90%
- Ⓚ 100%

8. Is this student currently receiving instruction in any of the following areas as part of a special education program? **(Grid in all that apply.)**

- Ⓐ Language development
- Ⓑ Reading
- Ⓒ Mathematics
- Ⓓ Speech (e.g., articulation, voice, speech flow)
- Ⓔ Self-control and deportment
- Ⓕ Personal care and basic life skills
- Ⓖ Vocational education
- Ⓗ Other (specify) ________________

Questions 9-22. What grade level of instruction is this student currently receiving in:

	Reading/Language Arts	Mathematics	Science
9. Lower than Kindergarten	Ⓐ	Ⓑ	Ⓒ
10. Kindergarten	Ⓐ	Ⓑ	Ⓒ
11. Grade 1	Ⓐ	Ⓑ	Ⓒ
12. Grade 2	Ⓐ	Ⓑ	Ⓒ
13. Grade 3	Ⓐ	Ⓑ	Ⓒ
14. Grade 4	Ⓐ	Ⓑ	Ⓒ
15. Grade 5	Ⓐ	Ⓑ	Ⓒ
16. Grade 6	Ⓐ	Ⓑ	Ⓒ
17. Grade 7	Ⓐ	Ⓑ	Ⓒ
18. Grade 8	Ⓐ	Ⓑ	Ⓒ
19. Grade 9	Ⓐ	Ⓑ	Ⓒ
20. Grade 10	Ⓐ	Ⓑ	Ⓒ
21. Grade 11	Ⓐ	Ⓑ	Ⓒ
22. Grade 12	Ⓐ	Ⓑ	Ⓒ

Questions 23-25. Is this student receiving essentially the same curriculum content as nondisabled students who are receiving instruction at the same grade level in:

23. Reading/Language Arts	Ⓐ No	Ⓑ Yes
24. Mathematics	Ⓐ No	Ⓑ Yes
25. Science	Ⓐ No	Ⓑ Yes

Questions 26-30. At what grade level is this student currently performing in:

	Reading/Language Arts	Mathematics	Science
26. Above grade level	Ⓐ	Ⓑ	Ⓒ
27. At grade level	Ⓐ	Ⓑ	Ⓒ
28. One year below grade level	Ⓐ	Ⓑ	Ⓒ
29. Two or more years below grade level	Ⓐ	Ⓑ	Ⓒ
30. I don't know.	Ⓐ	Ⓑ	Ⓒ

31. Are any accommodations or adaptations used for achievement testing for this student?

Ⓐ IEP states that student cannot be tested. [GO TO QUESTION 33.]

Ⓑ No [GO TO QUESTION 33.]

Ⓒ Yes

32. If the answer to question 31 is "Yes," which accommodations or adaptations are used for achievement testing for this student?

32a. Presentation Accommodation (Grid in all that apply.)

- Ⓐ Read directions aloud
- Ⓑ Read problems aloud (except on reading tests)
- Ⓒ Signing of directions
- Ⓓ Use of taped version of test
- Ⓔ Assistance with interpretation of directions
- Ⓕ Braille edition of test
- Ⓖ Large-print edition of test
- Ⓗ Use of magnifying equipment
- Ⓘ Other (specify) ________________

32b. Response Accommodation (Grid in all that apply.)

- Ⓐ Response in Braille
- Ⓑ Response in sign language
- Ⓒ Oral responses
- Ⓓ Pointing to answers
- Ⓔ Tape recording of answers
- Ⓕ Use of computer to respond
- Ⓖ Use of a typewriter to respond
- Ⓗ Use of a calculator including talking or Braille calculators
- Ⓘ Use of template to respond
- Ⓙ Use of a large marking pen or specially designed writing tool
- Ⓚ Other (specify) ________________

32c. Setting Accommodation (Grid in all that apply.)

- Ⓐ Test in small group
- Ⓑ Test individually
- Ⓒ Other (specify) ________________

32d. Timing Accommodation (Grid in all that apply.)

- Ⓐ Extended time
- Ⓑ More breaks during test
- Ⓒ Test sessions over several days
- Ⓓ Other (specify) ________________

33. In your judgment, could this student meaningfully participate in the NAEP assessment without accommodations or adaptations?

Ⓐ No

Ⓑ Yes

34. If accommodations and/or adaptations were available, how would this student participate in the NAEP assessment?

Ⓐ Without accommodations or adaptations

Ⓑ With the accommodations or adaptations specified for achievement testing of this student

Ⓒ The IEP team or an equivalent group has determined that the student cannot participate in assessments such as NAEP.

SECTION B: STUDENTS WITH LIMITED ENGLISH PROFICIENCY

Complete this section if the student has limited English proficiency.

35. How long has this student lived in the U.S.?

Ⓐ All his/her life

Ⓑ More than 5 years but not all his or her life

Ⓒ 3-5 years

Ⓓ Less than 3 years

Ⓔ I don't know.

36. What is this student's first or native language?

Ⓐ Spanish

Ⓑ Another language (specify) ____________________________

37. Since reaching school age, how regularly has this student attended school in the U.S. or in another country?

Ⓐ Continuously

Ⓑ Intermittently

Ⓒ Little or not at all

Ⓓ I don't know.

38. Counting this year, how many years has this student been enrolled in a school where English is the primary language of instruction?

Ⓐ The primary language of instruction in this school is not English

Ⓑ 1 year

Ⓒ 2 years

Ⓓ 3-5 years

Ⓔ More than 5 years

Ⓕ I don't know.

39. Counting this year, how many years has this student been receiving academic instruction (mathematics, reading/language arts, science) primarily in English?

- Ⓐ Student has been receiving instruction in his/her native language.
- Ⓑ 1 year
- Ⓒ 2 years
- Ⓓ 3-5 years
- Ⓔ More than 5 years
- Ⓕ I don't know.

40. Counting this year, how many years has this student received academic instruction specially designed for students with limited English proficiency (e.g., ESL, content based ESL, sheltered English content courses, native language support, native language instruction)?

- Ⓐ Student is not receiving instruction specially designed for LEP students [GO TO QUESTION 49.]
- Ⓑ 1 year
- Ⓒ 2 years
- Ⓓ 3 to 5 years
- Ⓔ More than 5 years

41. During those years while this student received specially designed academic instruction, in what language has instruction been provided?

- Ⓐ English only
- Ⓑ Primarily English with some instruction in first language
- Ⓒ About equally in English and in first language
- Ⓓ Primarily in first language with some instruction in English
- Ⓔ In first language only
- Ⓕ I don't know.

42. In which language could this student **best** demonstrate his/her knowledge of mathematics?

- (A) English
- (B) Spanish
- (C) Other (specify) ____________________

43. In which language could this student **best** demonstrate his/her knowledge of science?

- (A) English
- (B) Spanish
- (C) Other (specify) ____________________

Questions 44-45. During this school year, what percentage of this student's academic instruction is provided in English and what percentage in his/her native language? **(Grid in one letter for each language.)**

44. In English?

- (A) 0%
- (B) 10%
- (C) 20%
- (D) 30%
- (E) 40%
- (F) 50%
- (G) 60%
- (H) 70%
- (I) 80%
- (J) 90%
- (K) 100%

45. In his/her native language?

- (A) 0%
- (B) 10%
- (C) 20%
- (D) 30%
- (E) 40%
- (F) 50%
- (G) 60%
- (H) 70%
- (I) 80%
- (J) 90%
- (K) 100%

Questions 46-48. During this school year, has this student received any of the following types of instruction specially designed for LEP students in academic classes? **(Grid in one letter for each language.)**

	Specially designed instruction in English (such as ESL)	Native language instruction	Mainstreamed with no specially designed instruction
46. Reading/Language Arts	Ⓐ	Ⓑ	Ⓒ
47. Mathematics	Ⓐ	Ⓑ	Ⓒ
48. Science	Ⓐ	Ⓑ	Ⓒ

Questions 49-63. What grade level of instruction in the English language is this student currently receiving in:

	Reading/Language Arts	Mathematics	Science
49. Student is receiving instruction in his/her native language only.	Ⓐ	Ⓑ	Ⓒ
50. Lower than Kindergarten	Ⓐ	Ⓑ	Ⓒ
51. Kindergarten	Ⓐ	Ⓑ	Ⓒ
52. Grade 1	Ⓐ	Ⓑ	Ⓒ
53. Grade 2	Ⓐ	Ⓑ	Ⓒ
54. Grade 3	Ⓐ	Ⓑ	Ⓒ
55. Grade 4	Ⓐ	Ⓑ	Ⓒ
56. Grade 5	Ⓐ	Ⓑ	Ⓒ
57. Grade 6	Ⓐ	Ⓑ	Ⓒ
58. Grade 7	Ⓐ	Ⓑ	Ⓒ
59. Grade 8	Ⓐ	Ⓑ	Ⓒ
60. Grade 9	Ⓐ	Ⓑ	Ⓒ
61. Grade 10	Ⓐ	Ⓑ	Ⓒ
62. Grade 11	Ⓐ	Ⓑ	Ⓒ
63. Grade 12	Ⓐ	Ⓑ	Ⓒ

Questions 64-66. Is this student receiving essentially the same curricular content in the English language as English-speaking students who are receiving instruction at the same grade level in:

	No	Yes	Student is receiving instruction in his/her native language only
64. Reading/Language Arts	Ⓐ	Ⓑ	Ⓒ
65. Mathematics	Ⓐ	Ⓑ	Ⓒ
66. Science	Ⓐ	Ⓑ	Ⓒ

Questions 67-72. At what grade level is this student currently performing in the English language in:

	Reading/Language Arts	Mathematics	Science
67. Student is receiving instruction in his/her native language only.	Ⓐ	Ⓑ	Ⓒ
68. Above grade level	Ⓐ	Ⓑ	Ⓒ
69. At grade level	Ⓐ	Ⓑ	Ⓒ
70. One year below grade level	Ⓐ	Ⓑ	Ⓒ
71. Two or more years below grade level	Ⓐ	Ⓑ	Ⓒ
72. I don't know.	Ⓐ	Ⓑ	Ⓒ

Questions 73-78. At what grade level is this student currently performing in his/her native language in:

	Reading/Language Arts	Mathematics	Science
73. Student is receiving instruction in his/her native language only.	A	B	C
74. Above grade level	A	B	C
75. At grade level	A	B	C
76. One year below grade level	A	B	C
77. Two or more years below grade level	A	B	C
78. I don't know.	A	B	C

Questions 79-82. How would you characterize this student's English proficiency compared to native English speakers? **(Grid in one letter on each line.)**

	Good (LEP advanced)	Fair (LEP intermediate)	Poor (LEP beginning)	No proficiency	I don't know
79. Understanding	A	B	C	D	E
80. Speaking	A	B	C	D	E
81. Reading	A	B	C	D	E
82. Writing	A	B	C	D	E

Questions 83-84. How would you characterize this student's native language proficiency for his/her age? **(Grid in one letter on each line.)**

	Excellent	Good	Fair	Poor	No proficiency	I don't know
83. Reading	A	B	C	D	E	F
84. Writing	A	B	C	D	E	F

85. Are any accommodations or adaptations used for achievement testing for this student?

- Ⓐ No [GO TO QUESTION 87.]
- Ⓑ Yes

86. If yes, which accommodations or adaptations are used for achievement testing for this student? **(Grid in all that apply.)**

- Ⓐ Native language version of test
- Ⓑ Word lists or glossaries
- Ⓒ English/native language dictionary
- Ⓓ Help from a native speaker in interpreting directions and questions
- Ⓔ Directions read aloud twice in English
- Ⓕ Questions read aloud in English
- Ⓖ Extended time
- Ⓗ Other (specify) ____________________________

87. In your judgment, could this student meaningfully participate in the NAEP assessment without accommodations?

- Ⓐ No
- Ⓑ Yes

88. If accommodations and/or adaptations were available, how would this student participate in the NAEP assessment?

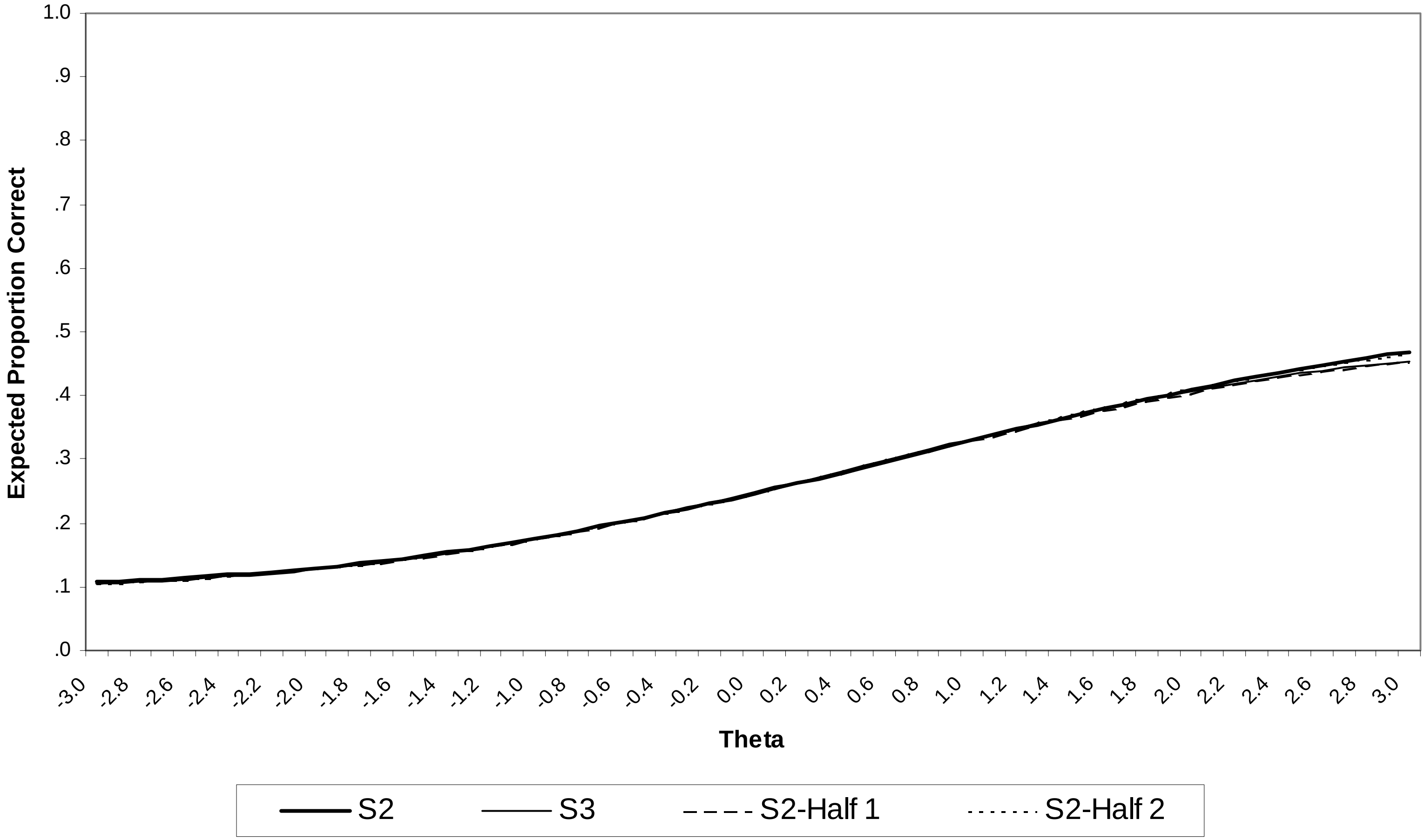
- Ⓐ In English without accommodations or adaptations
- Ⓑ In English with the accommodations or adaptations specified for achievement testing of this student
- Ⓒ In his/her native language
- Ⓓ In his/her native language with the accommodations or adaptations specified for achievement testing of this student
- Ⓔ The student would not participate

THANK YOU FOR YOUR COOPERATION.

Appendix C

Test Characteristic Curves for Science, Grades 8 & 12

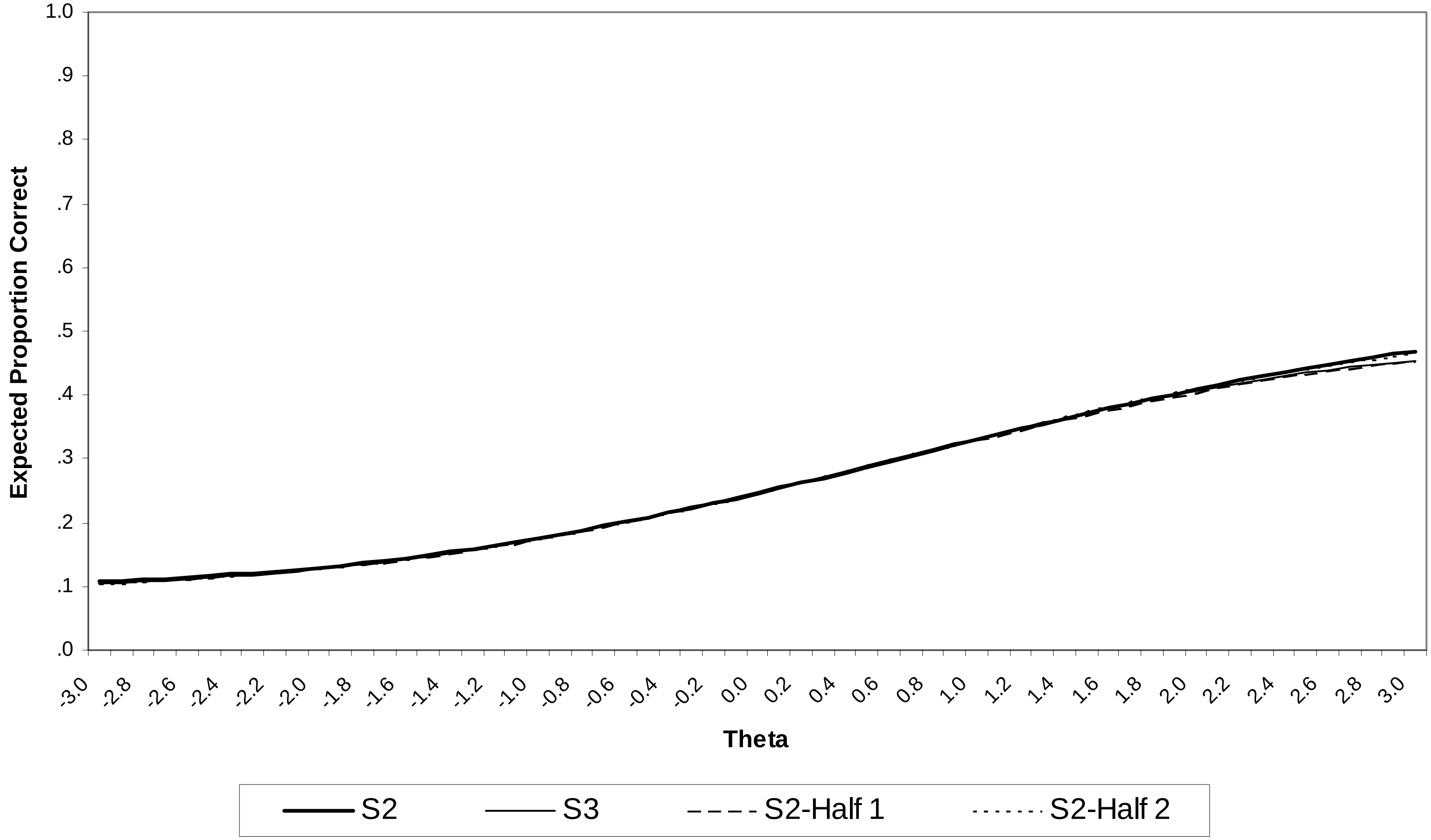
Figure C.1 — Test characteristic curves for grade-8 physical science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figure C.2 — Test characteristic curves for grade-8 earth science scale

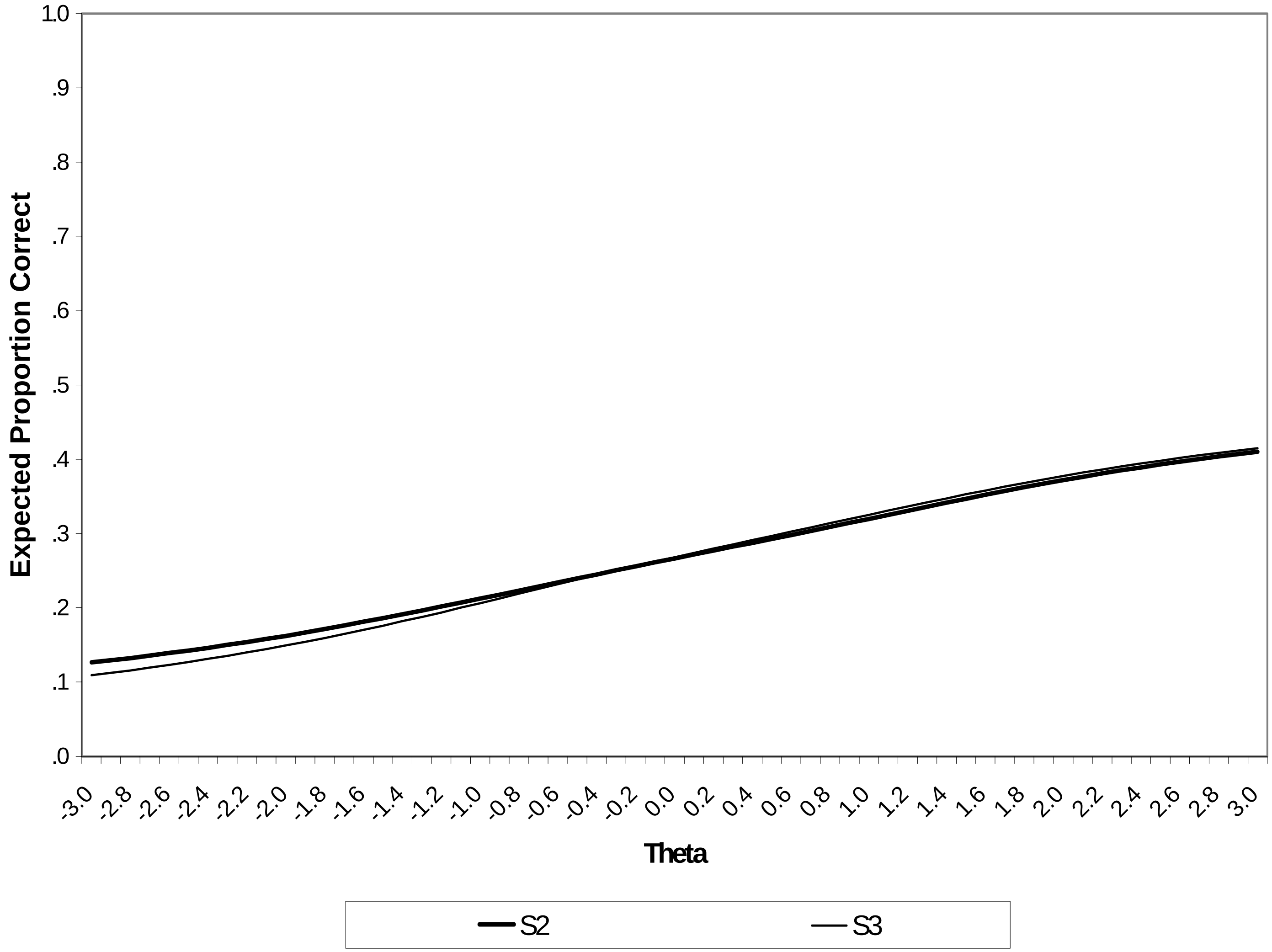
THE NATION'S REPORT CARD naep



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

THE NATION'S REPORT CARD naep

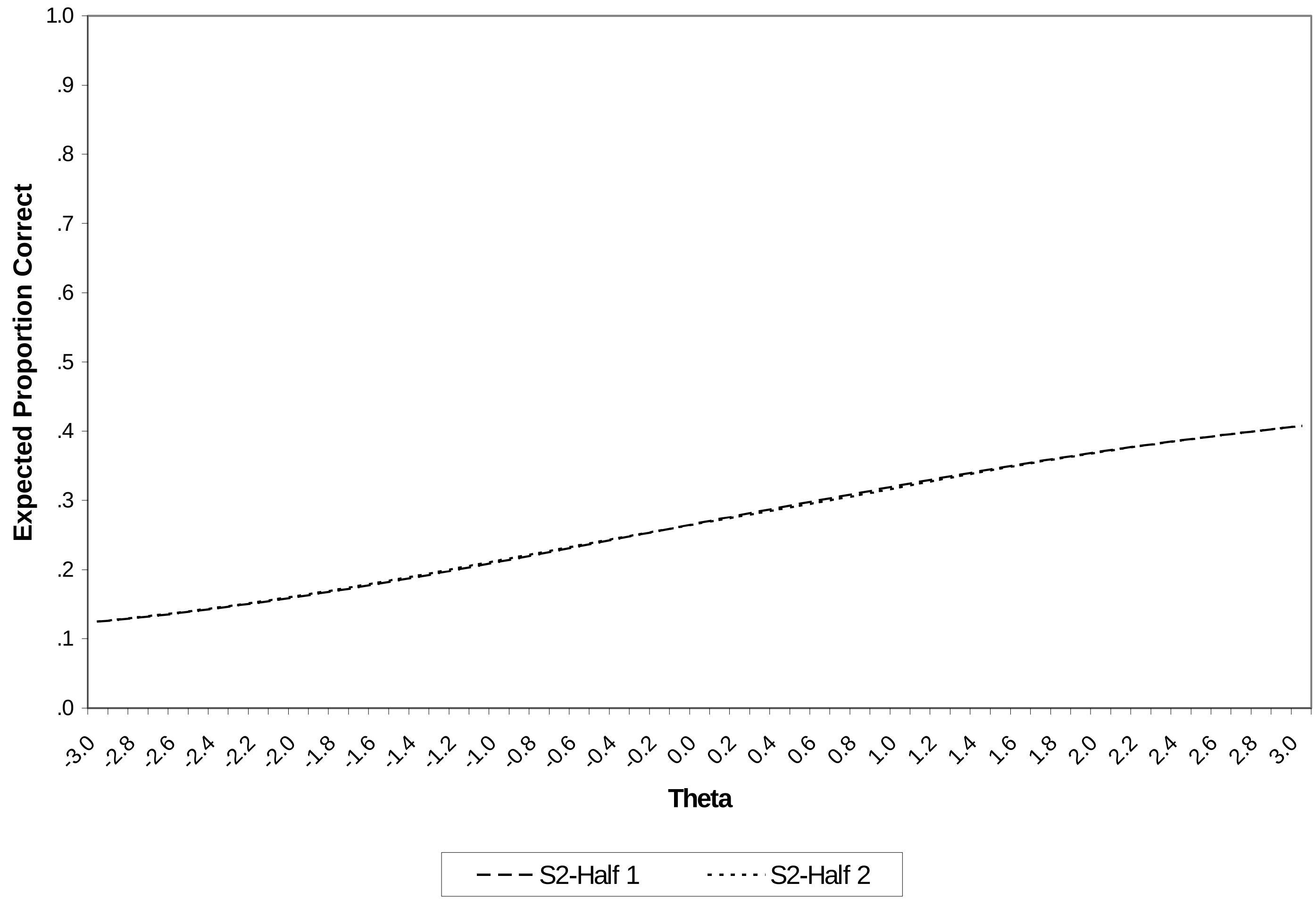
Figure C.3 — Test characteristic curves for grade-8 life science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

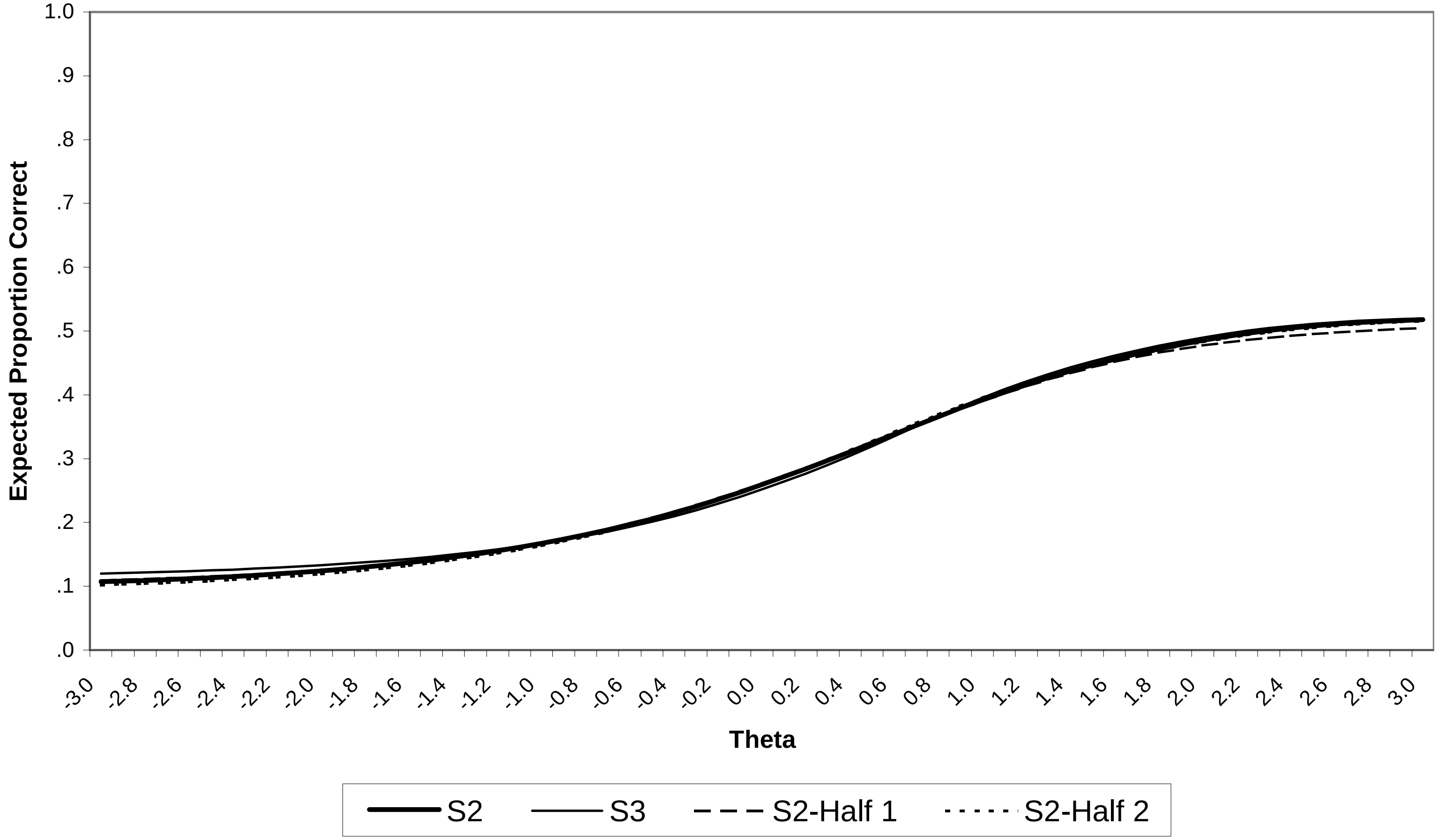
Figure C.4 — Test characteristic curves for grade-8 life science scale – S2 random halves

THE NATION'S REPORT CARD



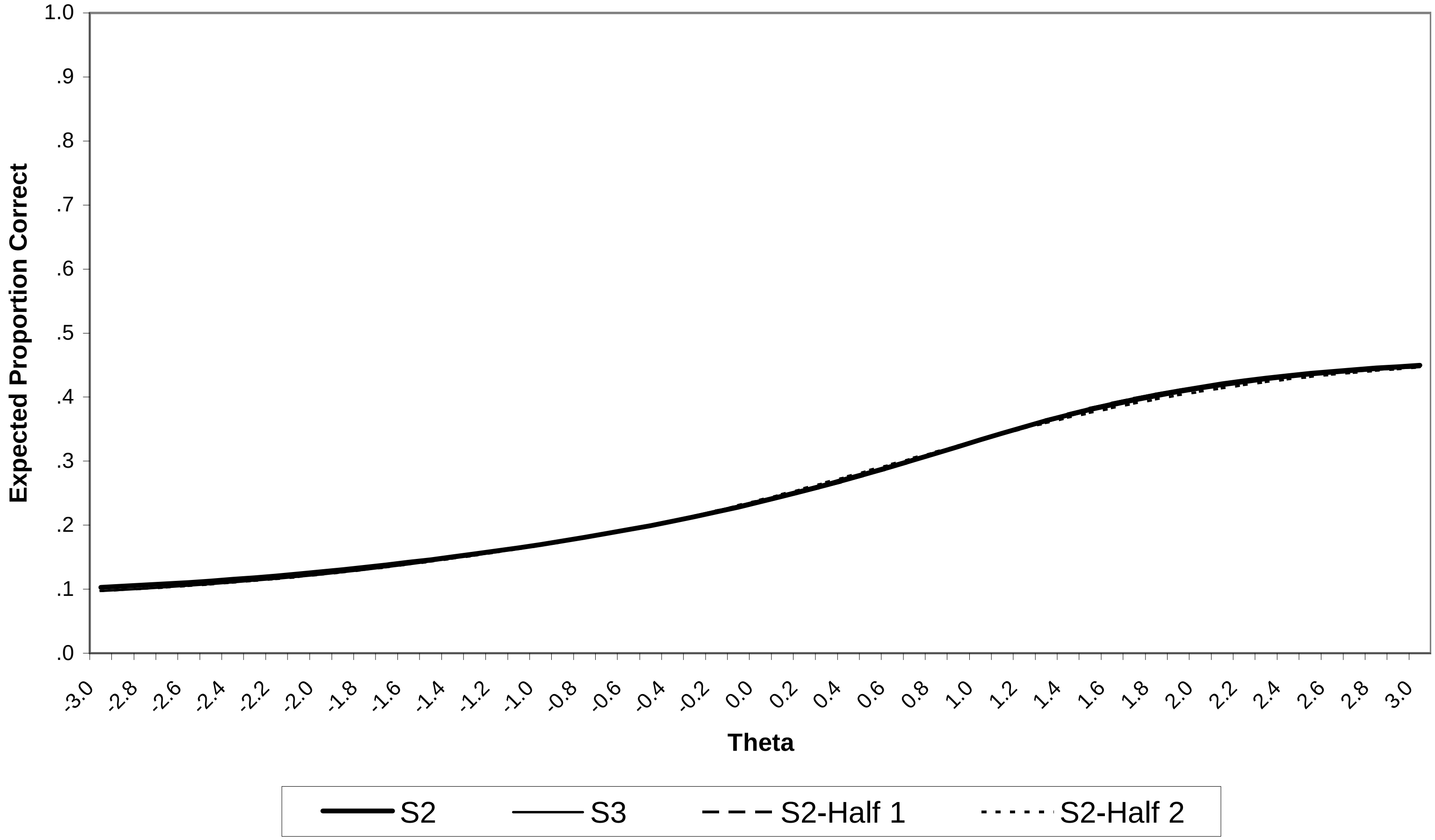
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figure C.5 — Test characteristic curves for grade-12 physical science scale



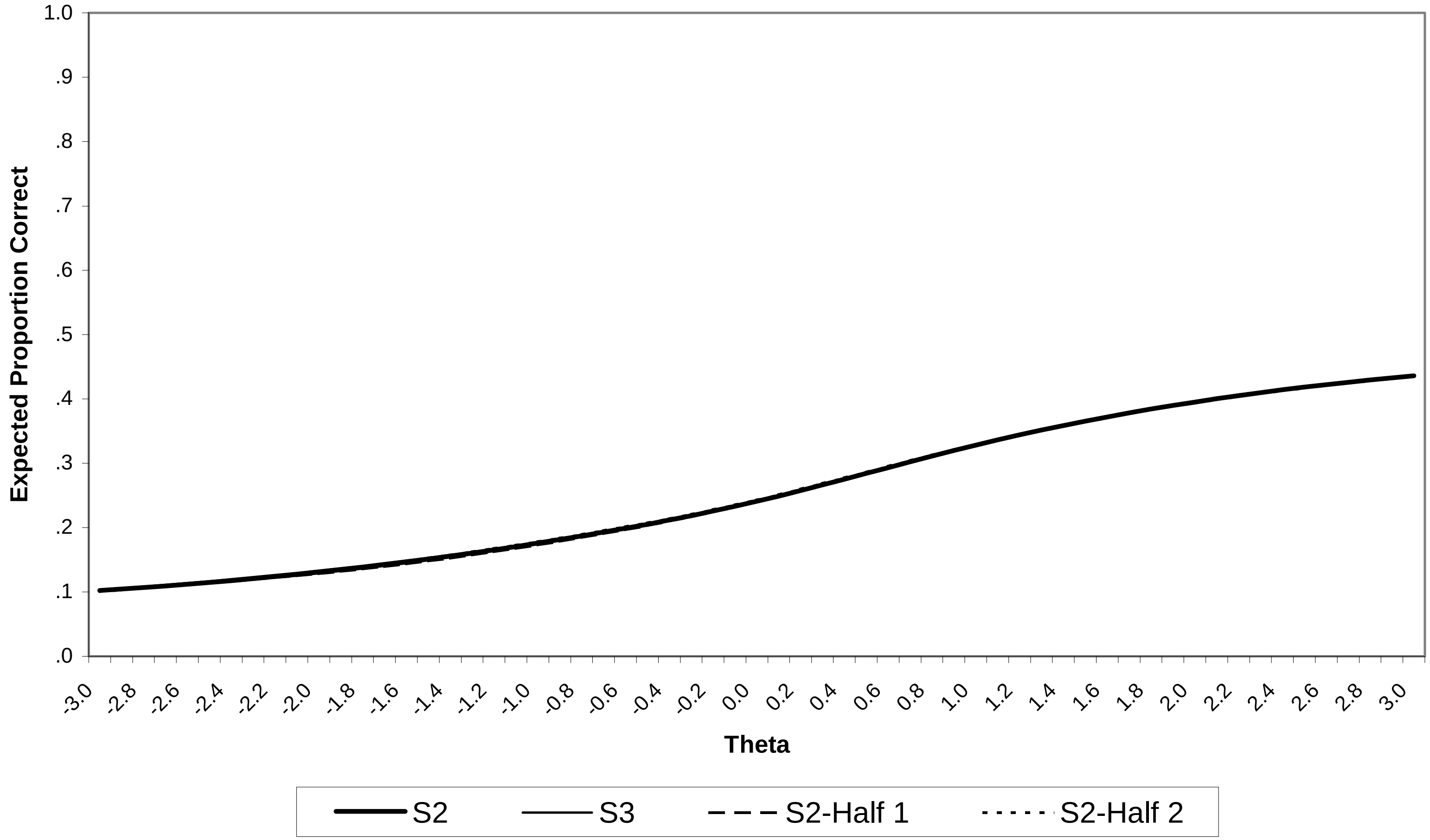
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figure C.6 — Test characteristic curves for grade-12 earth science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figure C.7— Test characteristic curves for grade-12 life science scale

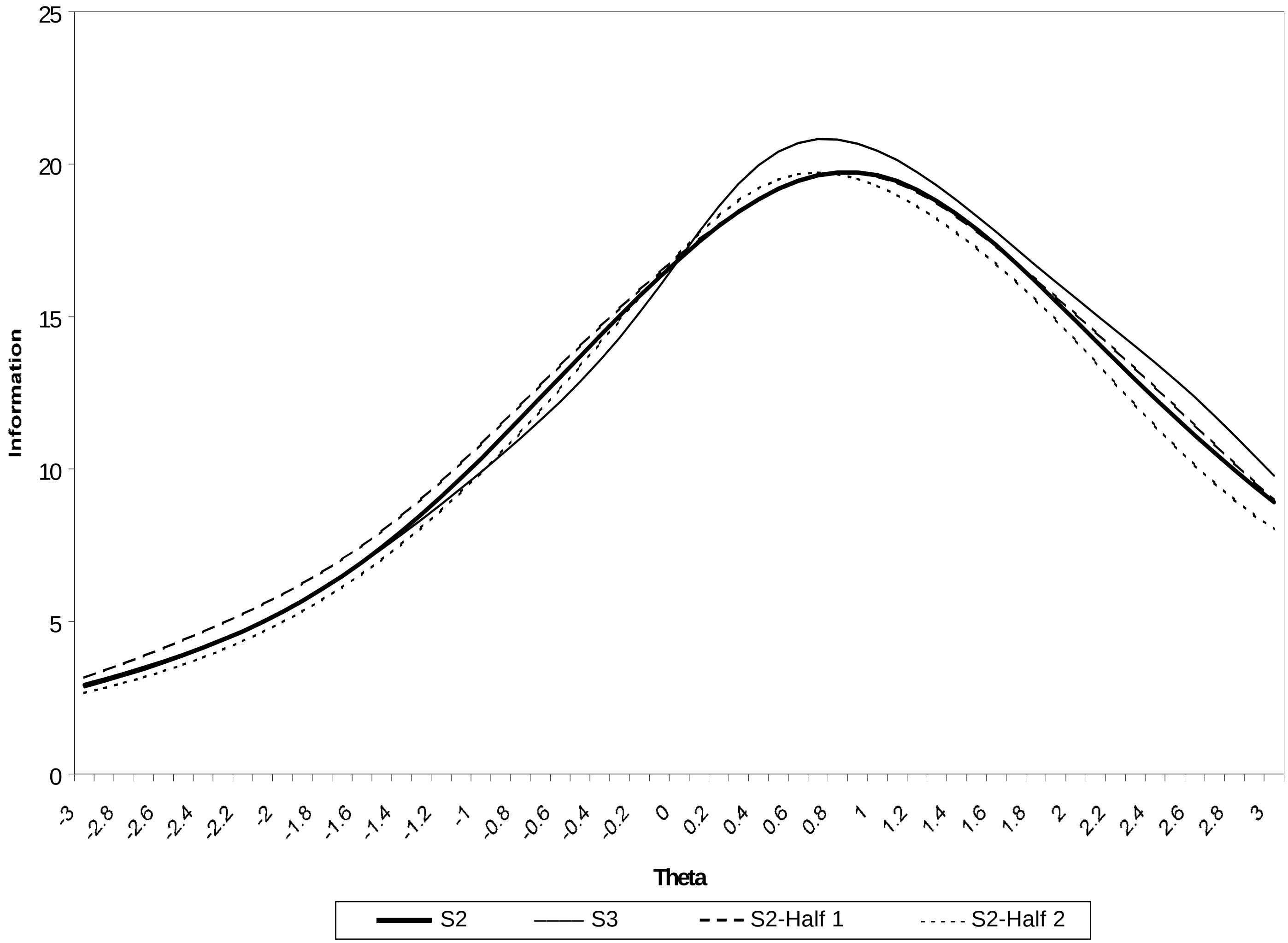


SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Appendix D

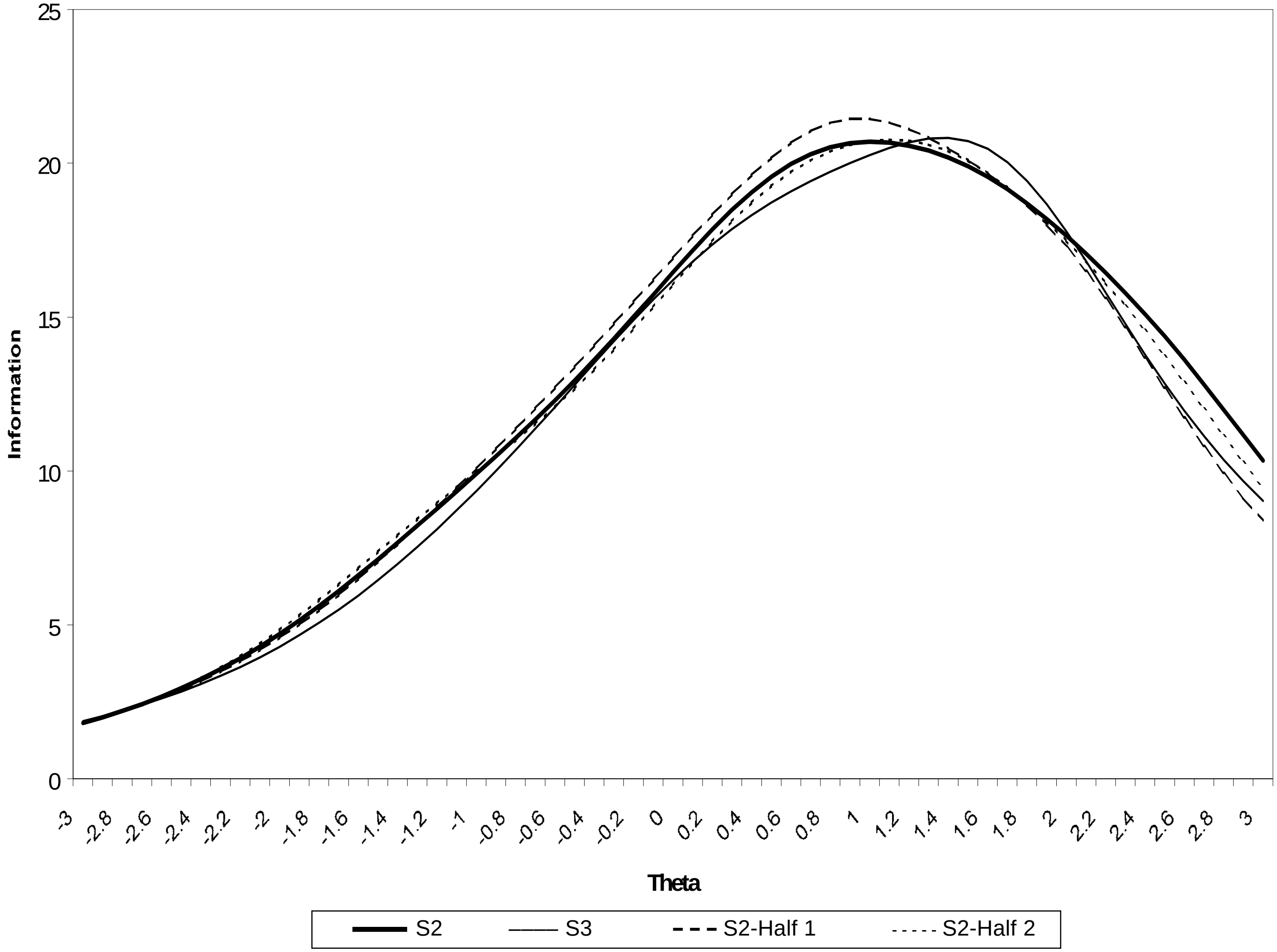
Test Information Curves for Science, Grades 8 & 12

Figure D.1 — Test information curves for grade-8 physical science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

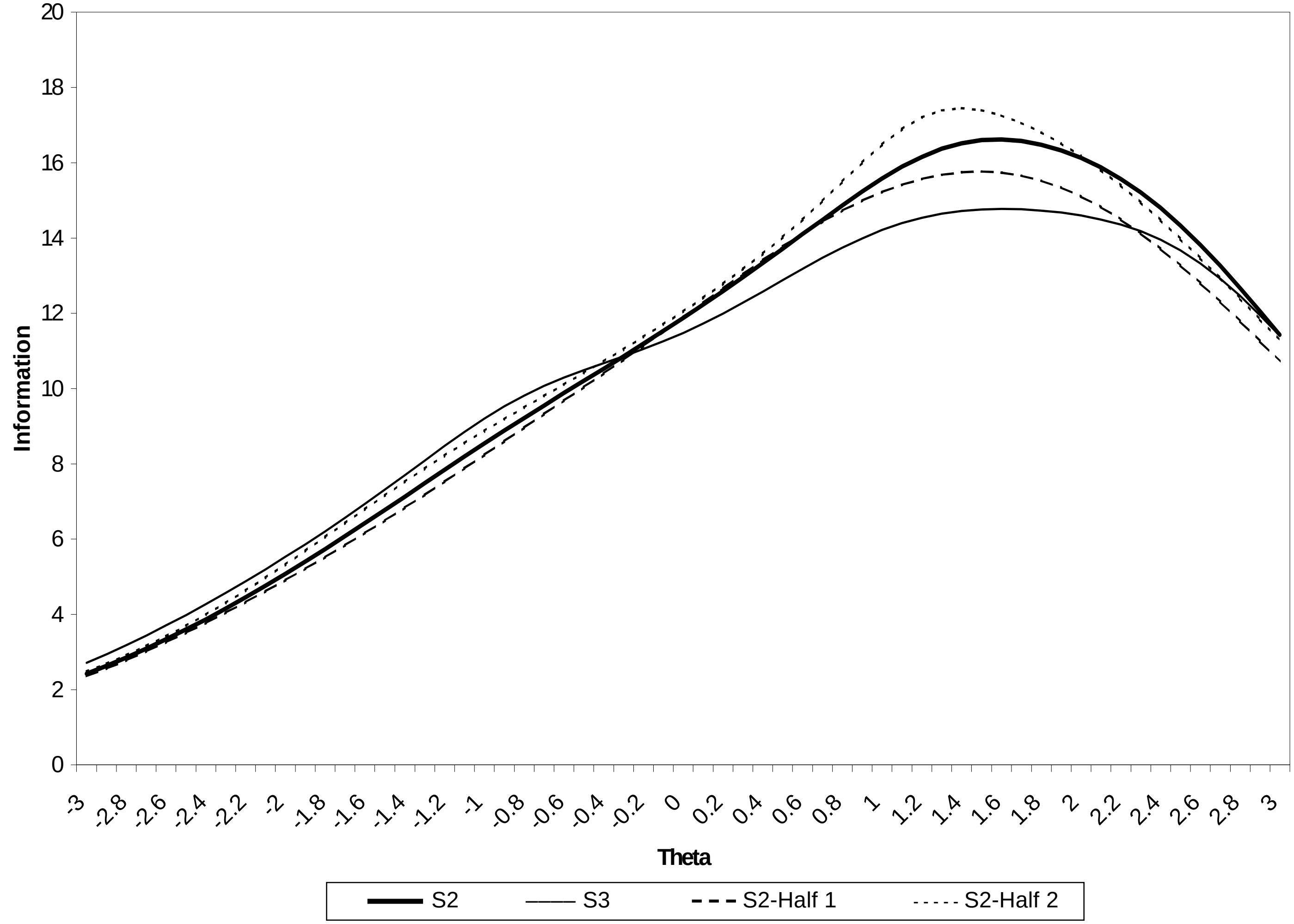
Figure D.2 — Test information curves for grade-8 earth science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

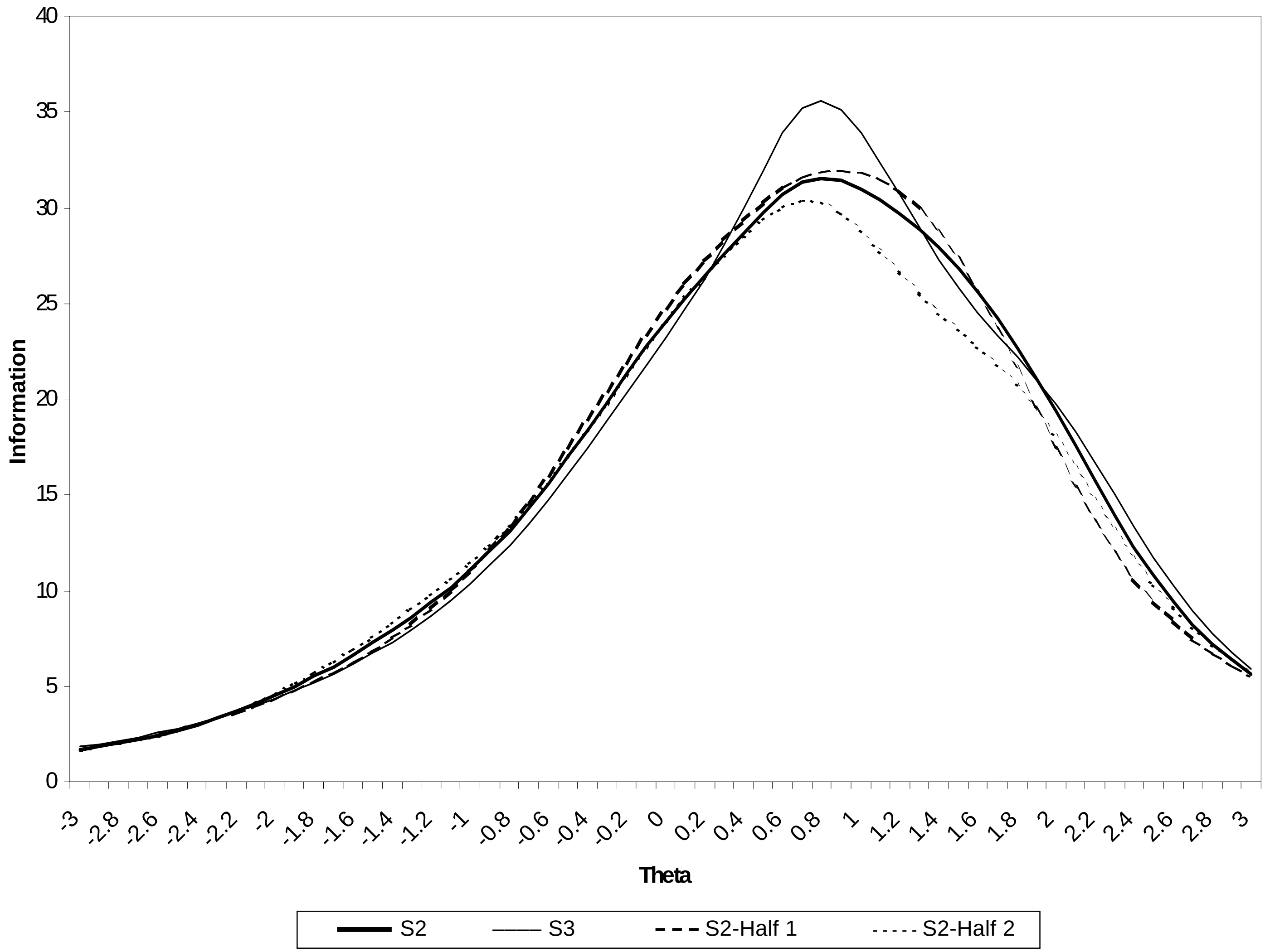
THE NATION'S REPORT CARD

Figure D.3 — Test information curves for grade-8 life science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

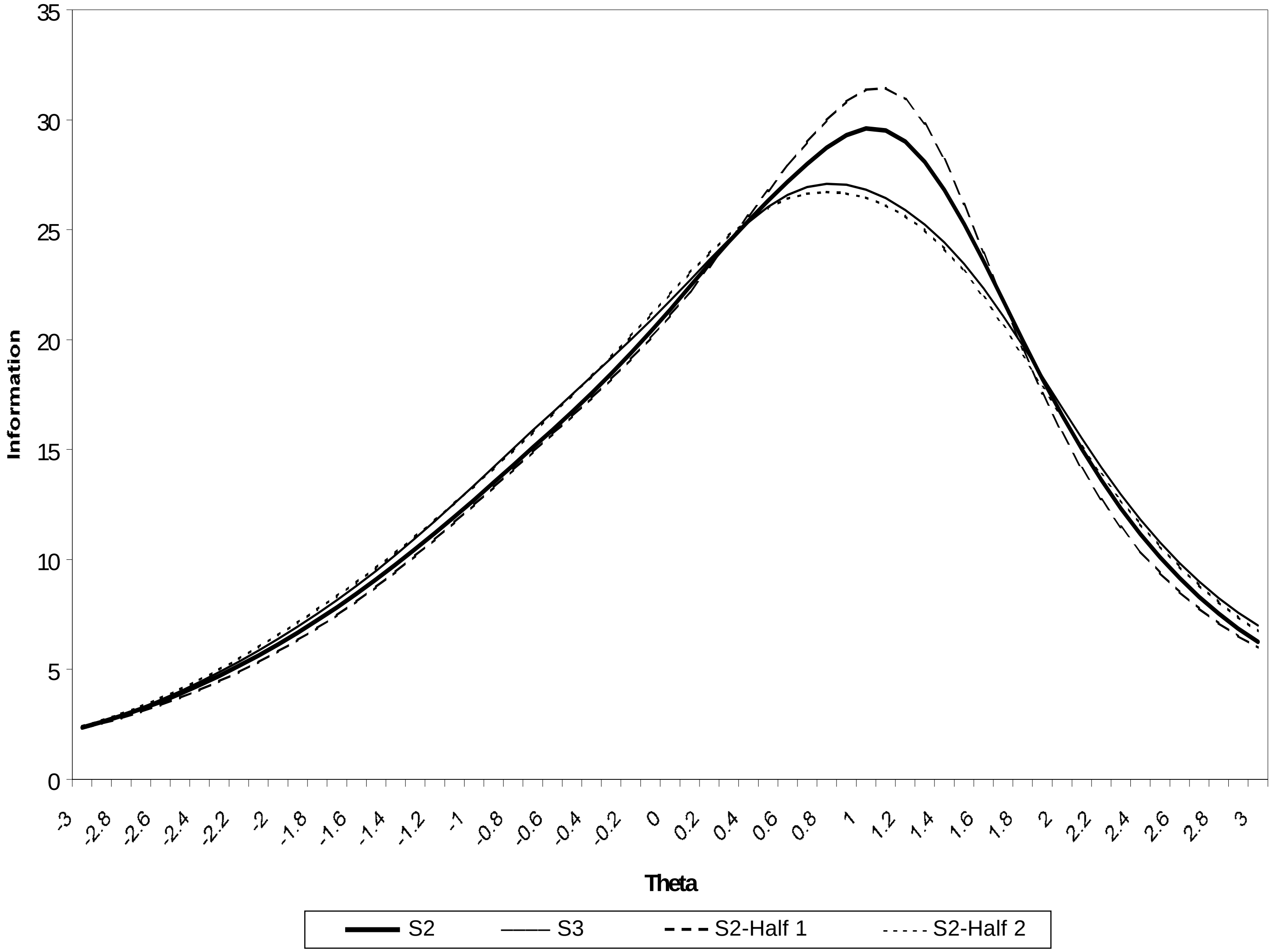
Figure D.4 — Test information curves for grade-12 physical science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

THE NATION'S REPORT CARD naep

Figure D.5 — Test information curves for grade-12 earth science scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

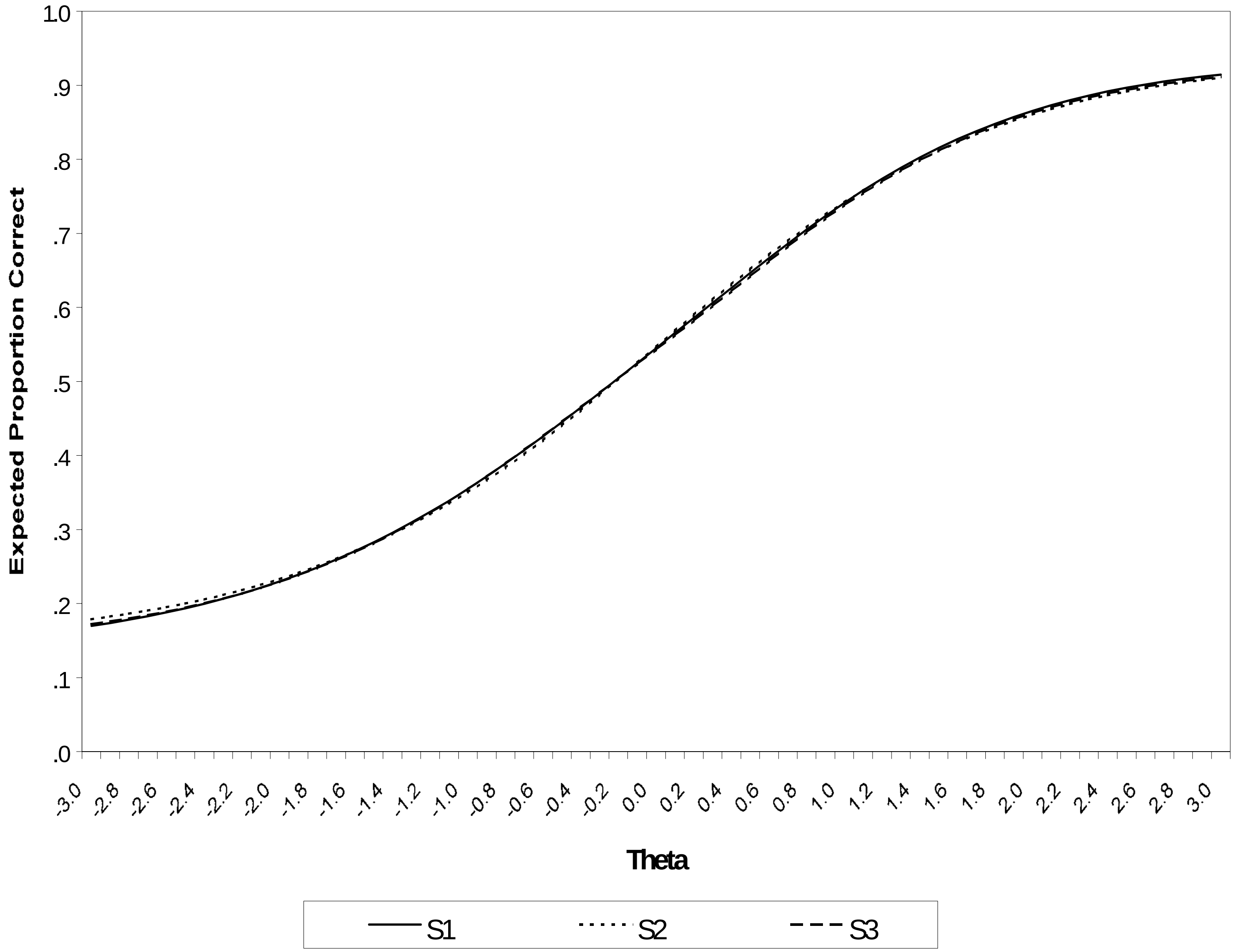
Figure D.6 — Test information curves for grade-12 life science scale

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Appendix E

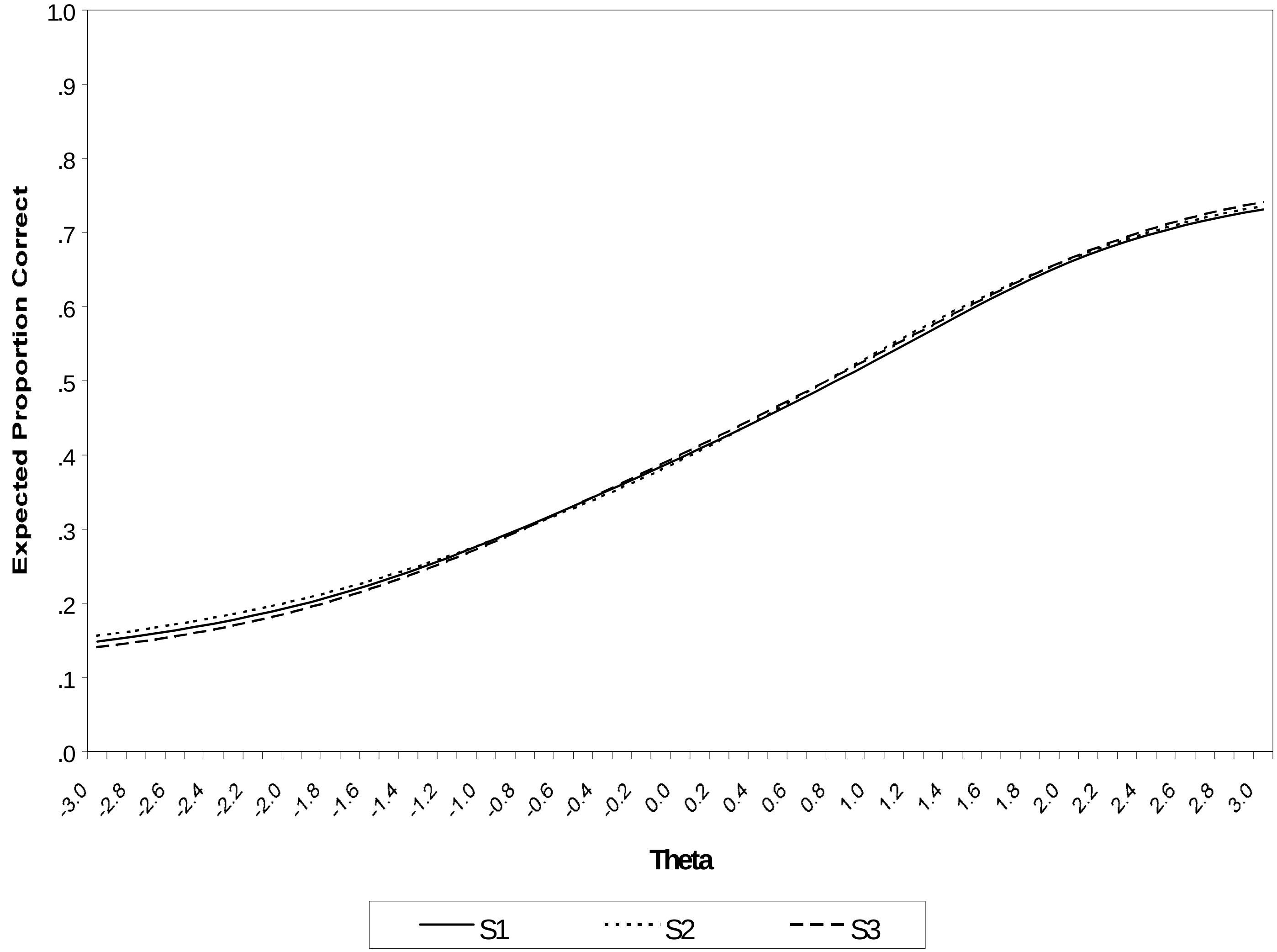
Test Characteristic Curves for Mathematics, Grades 8 & 12

Figure E.1 — Test characteristic curves for grade-8 numbers and operations scale



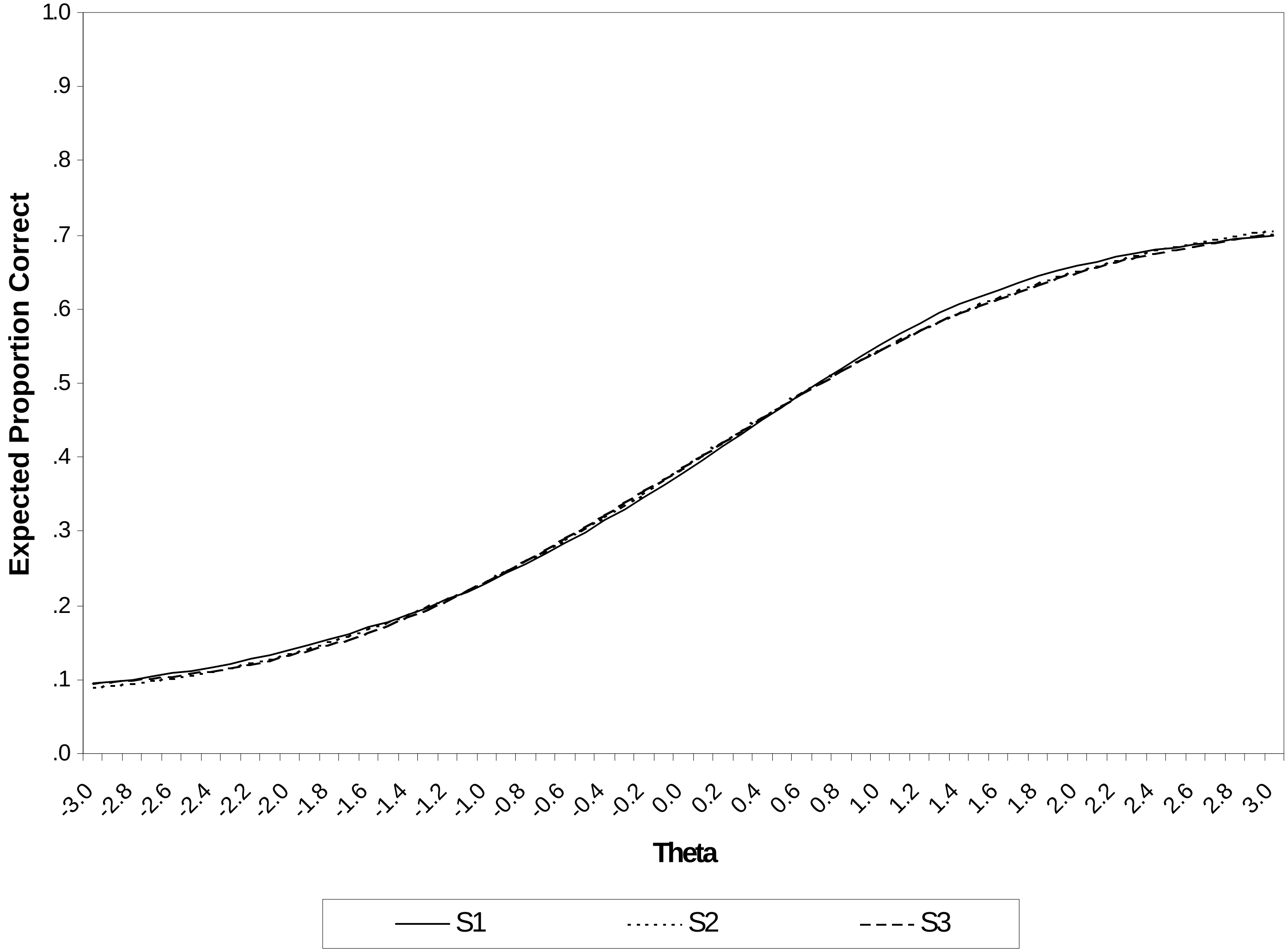
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure E.2 — Test characteristic curves for grade-8 measurement scale



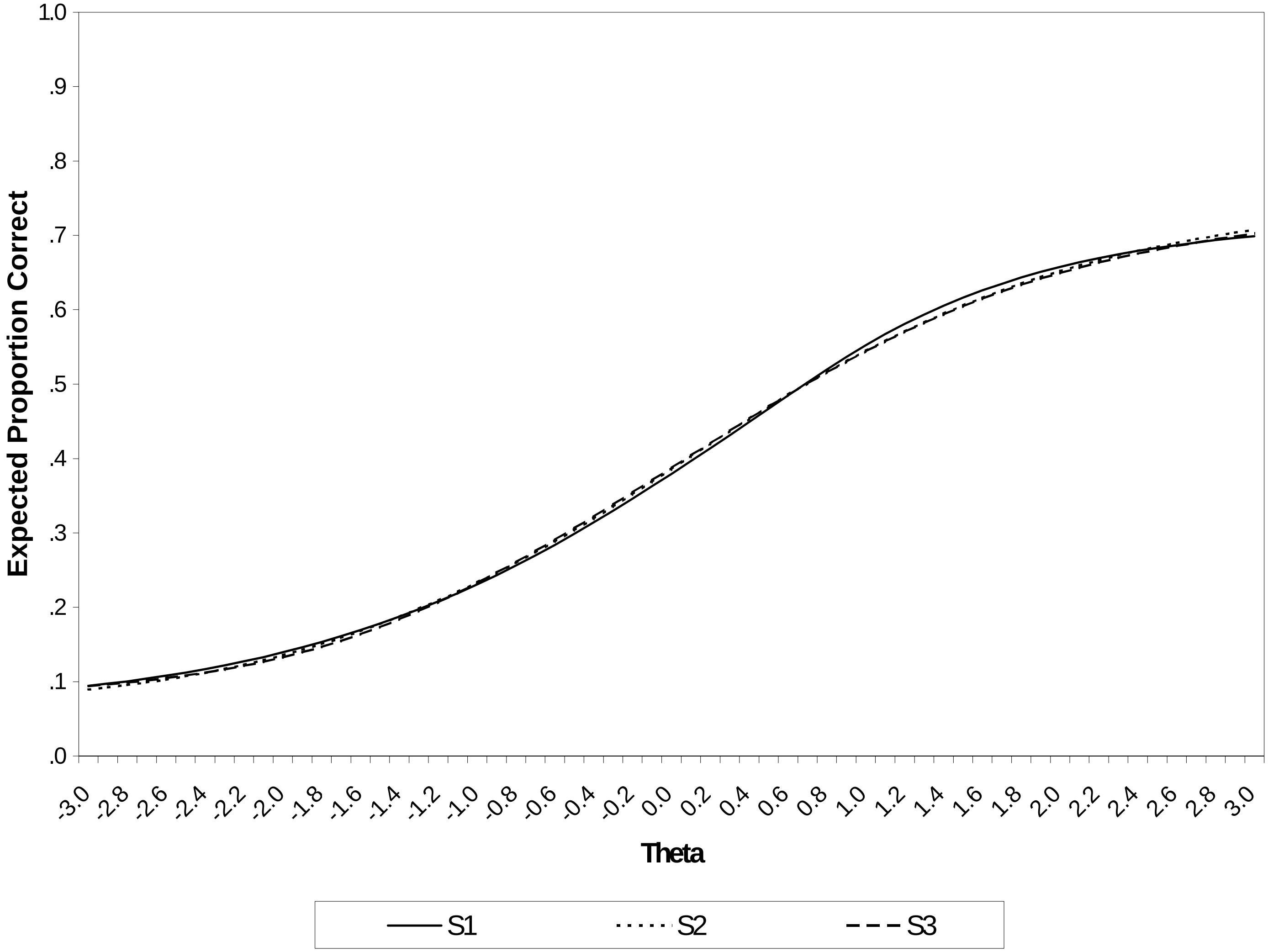
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure E.3 — Test characteristic curves for grade-8 geometry scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

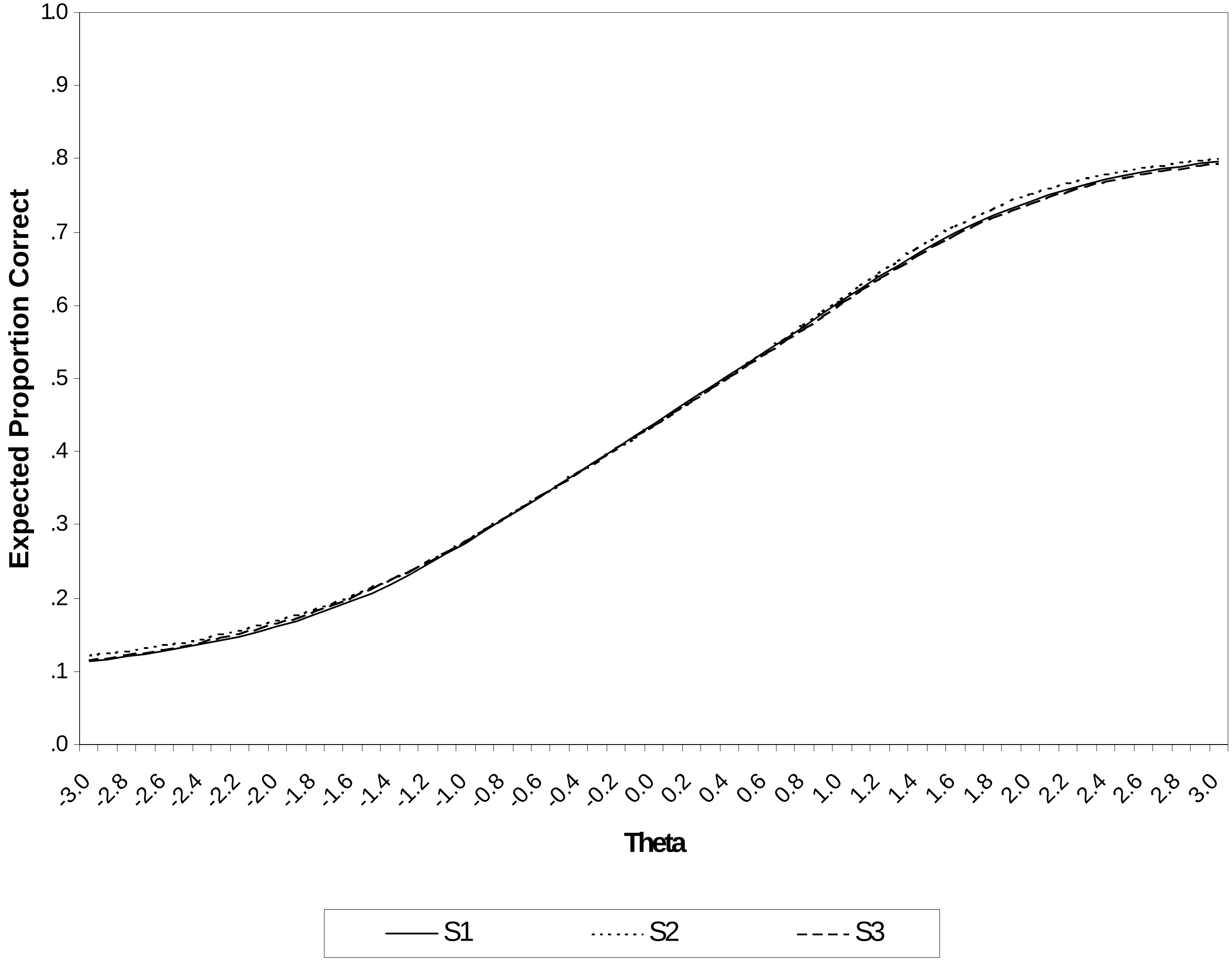
Figure E.4 — Test characteristic curves for grade-8 data analysis scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

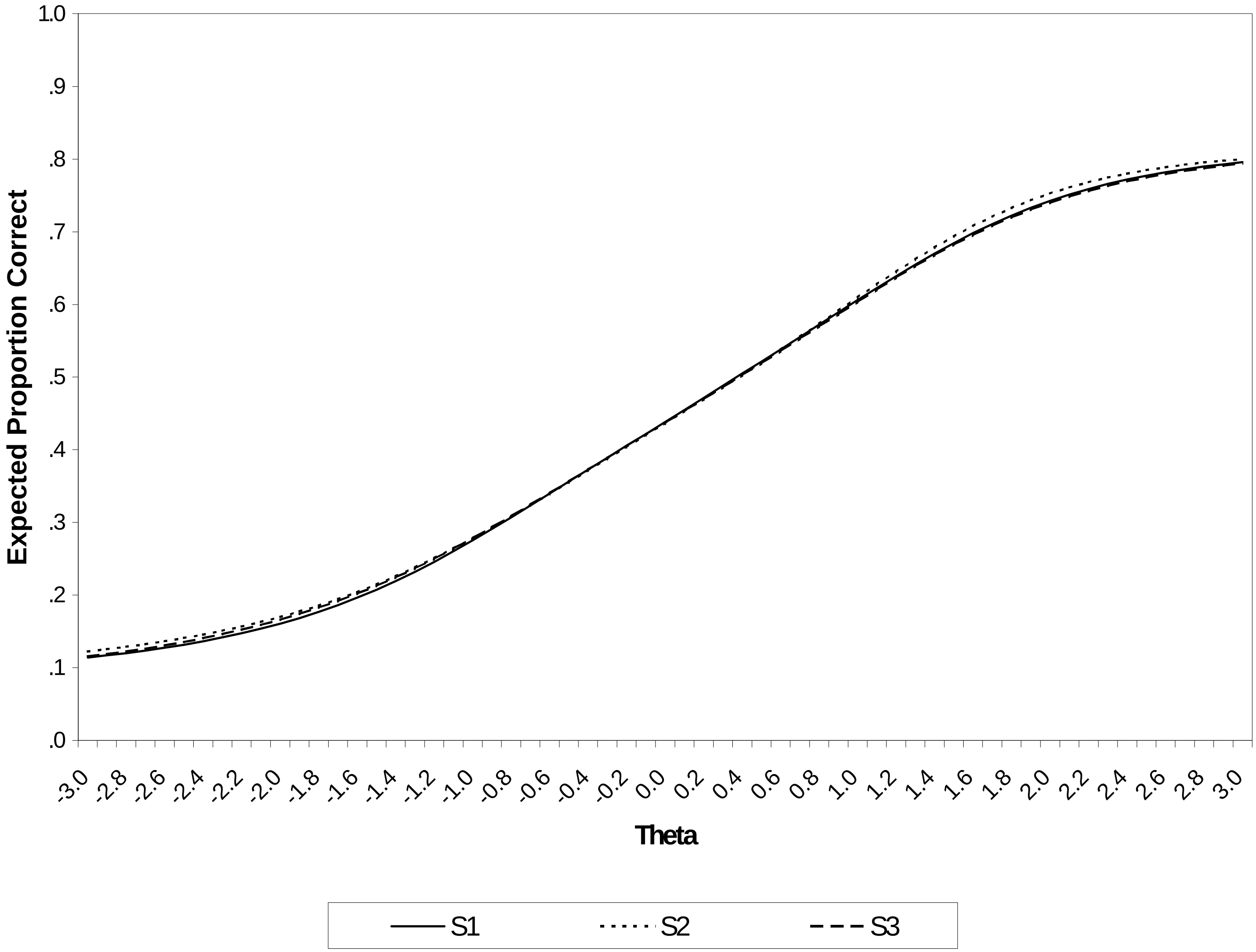
THE NATION'S REPORT CARD naep

Figure E.5 — Test characteristic curves for grade-12 numbers and operations scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

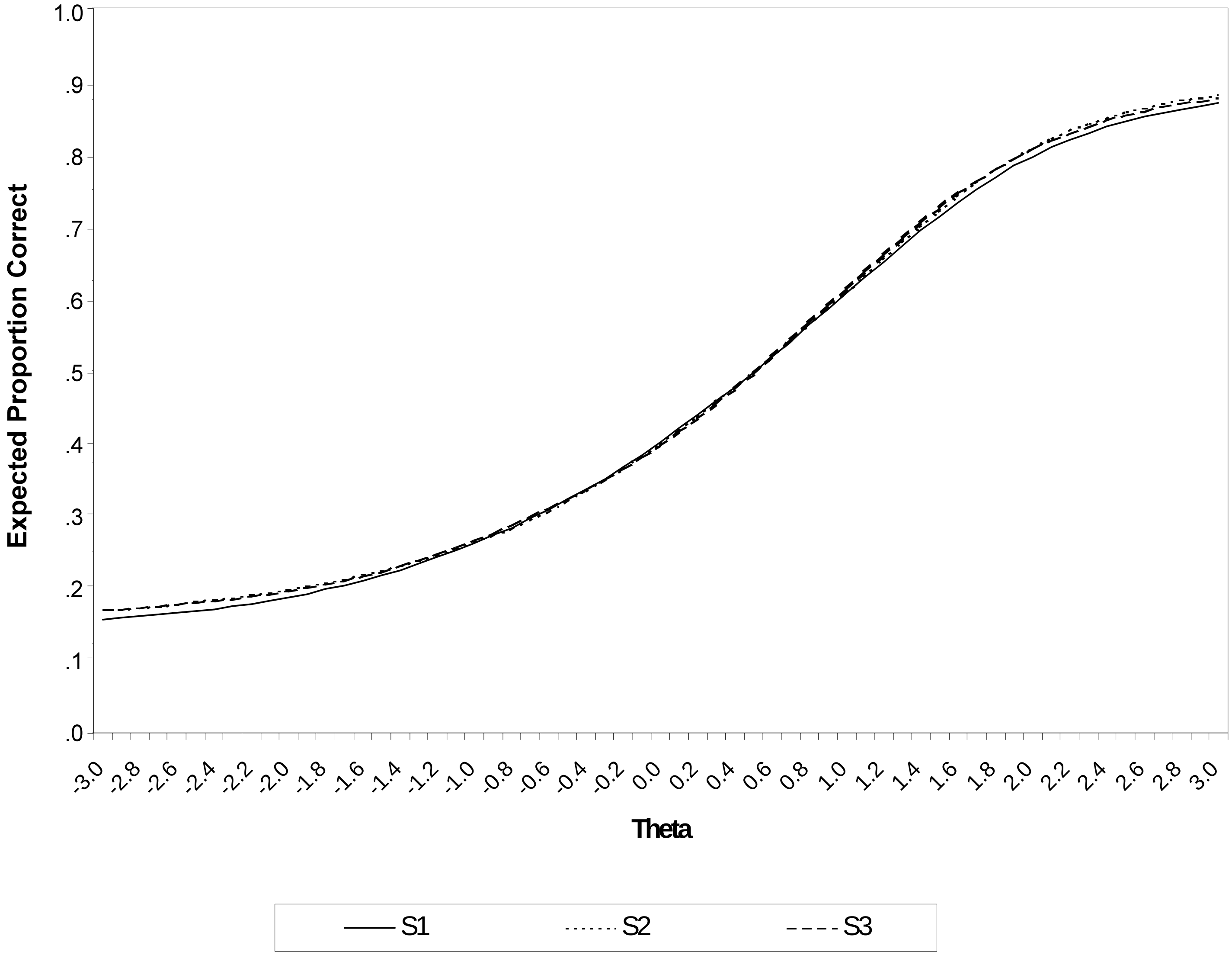
Figure E.6 — Test characteristic curves for grade-12 geometry scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

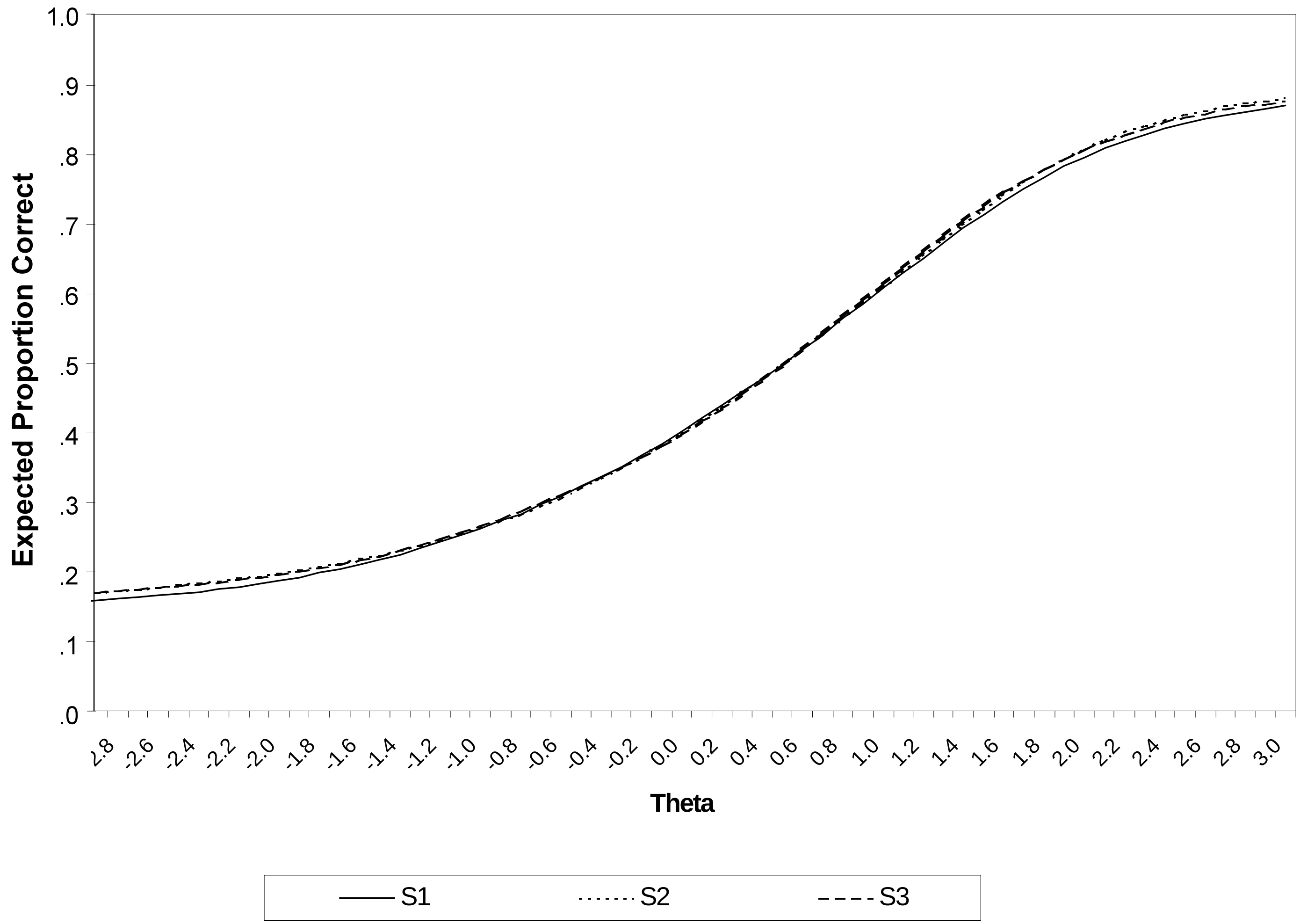
THE NATION'S REPORT CARD

Figure E.7 — Test characteristic curves for grade-12 data analysis scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure E.8 — Test characteristic curves for grade-12 algebra and functions scale

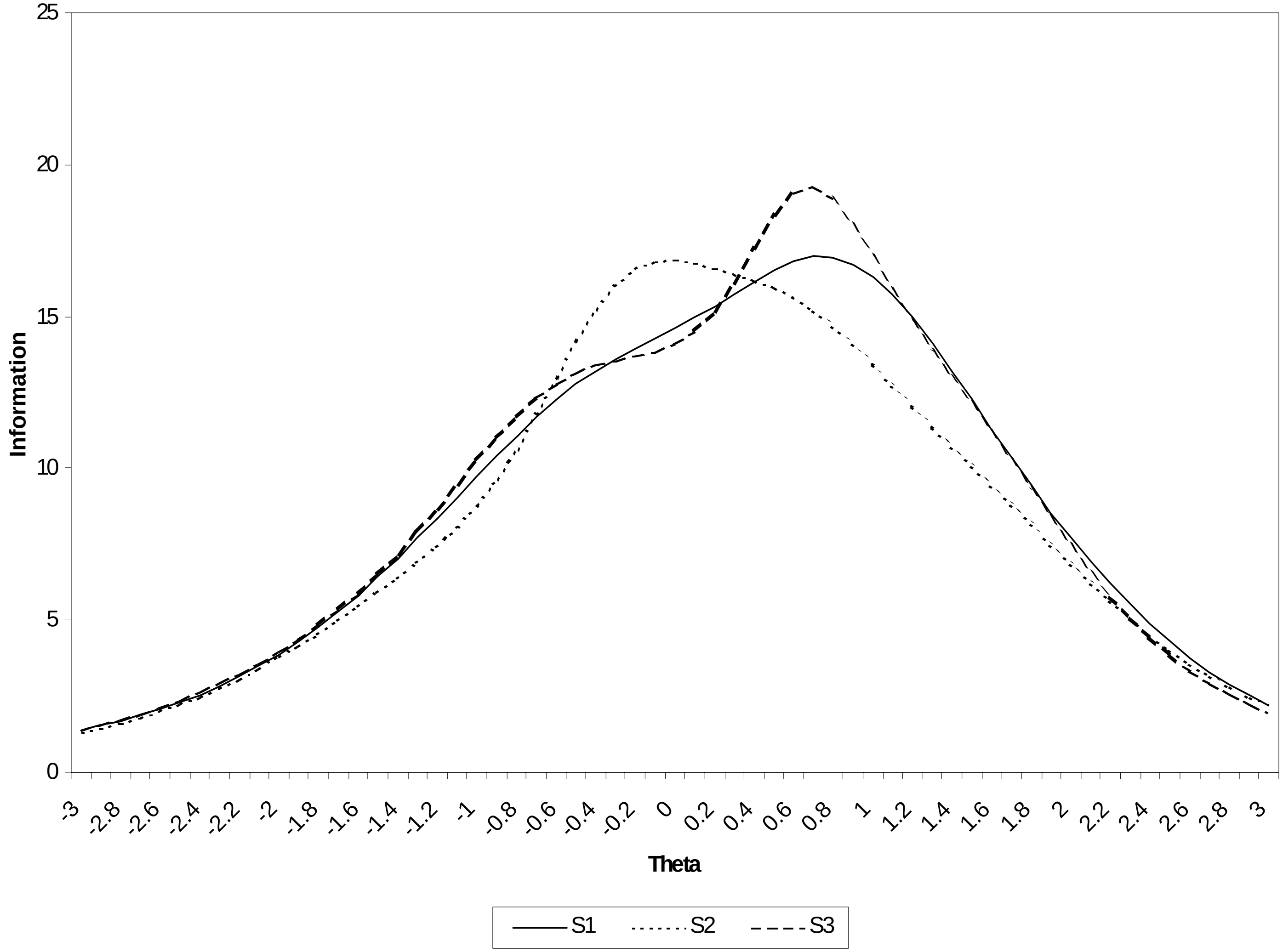


SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Appendix F

Test Information Curves for Mathematics, Grades 8 and 12

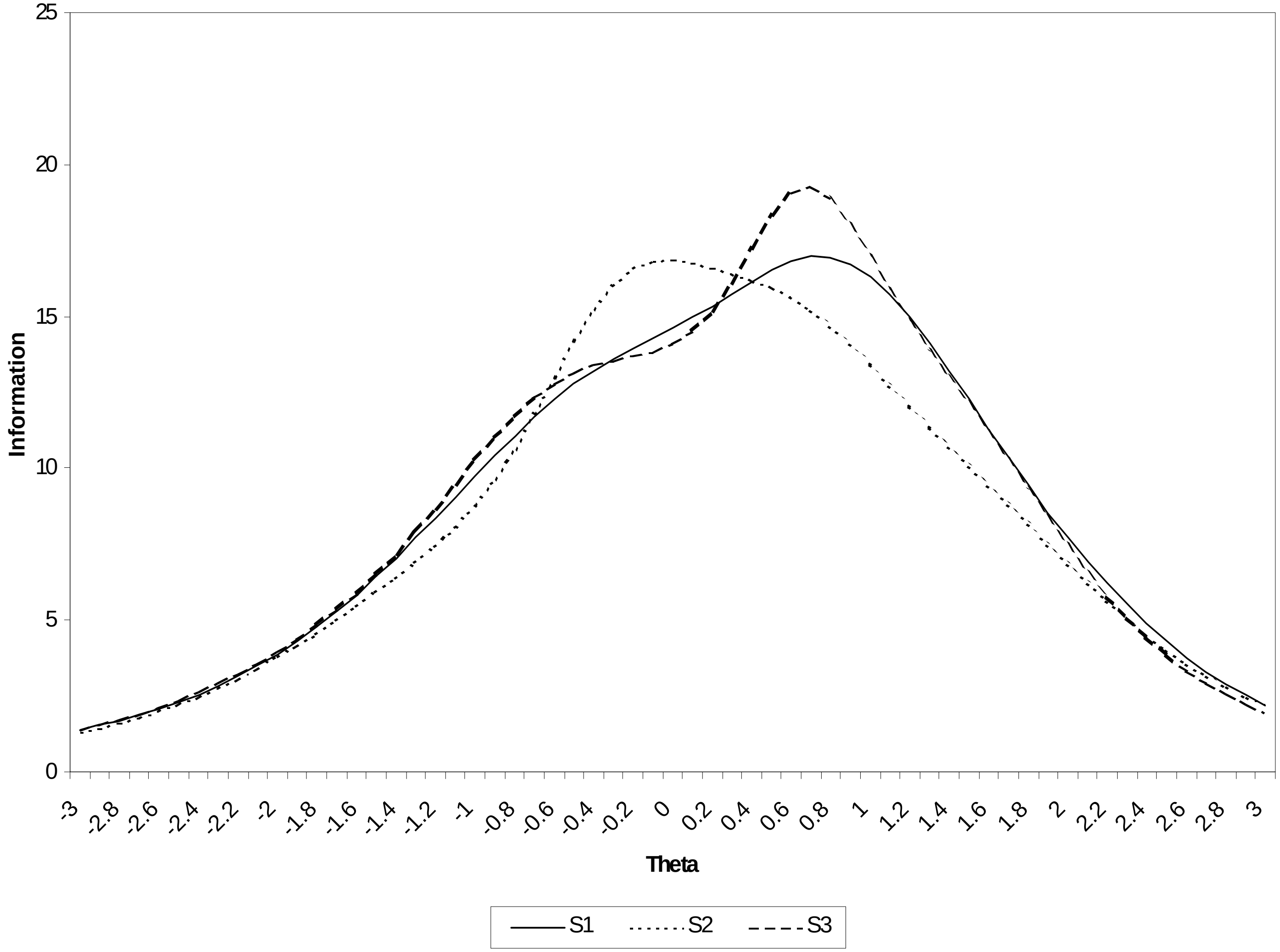
Figure F.1— Test information curves for grade-8 numbers and operations scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

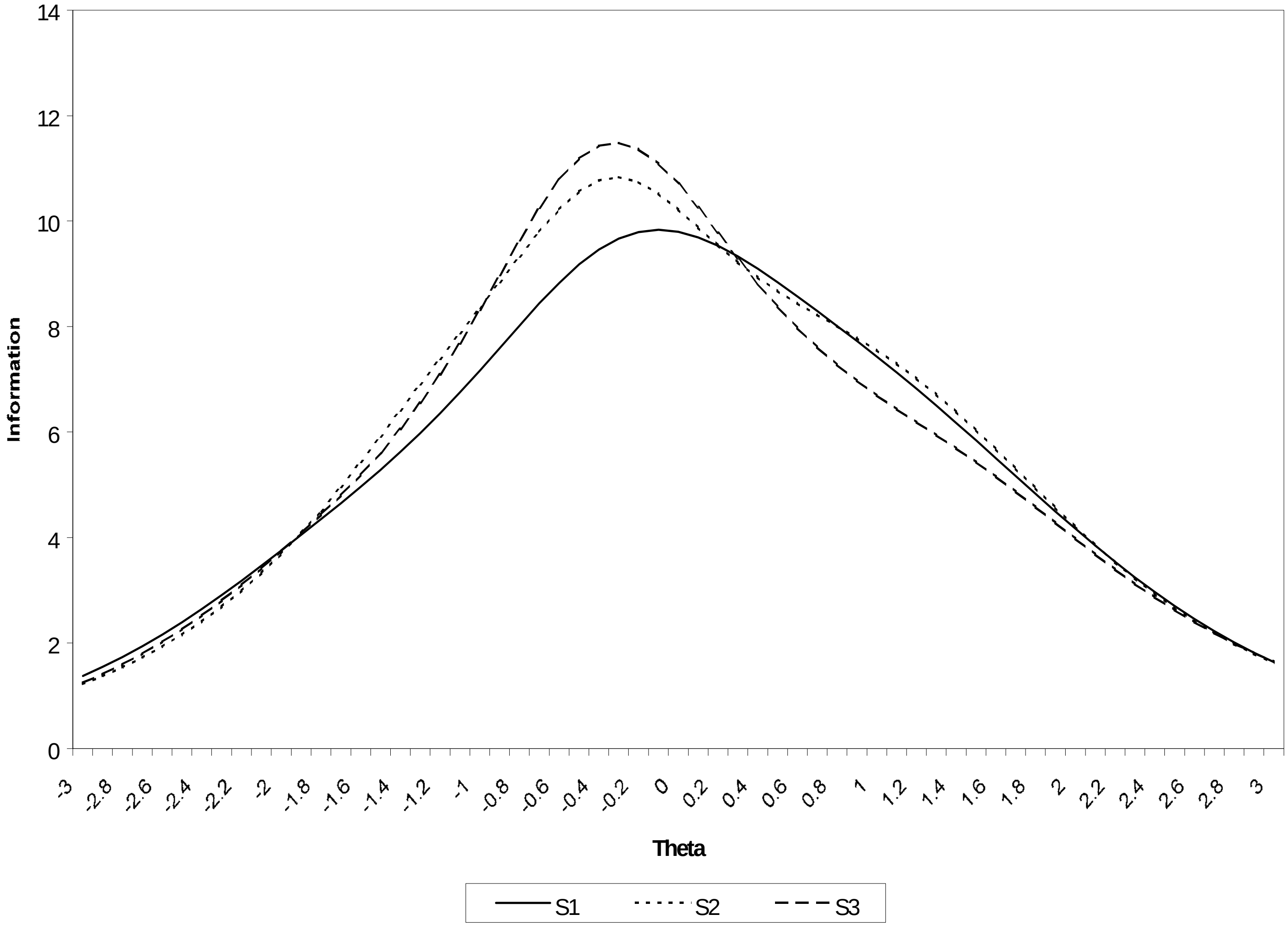
Figure F.2— Test information curves for grade-8 measurement scale

THE NATION'S REPORT CARD naep



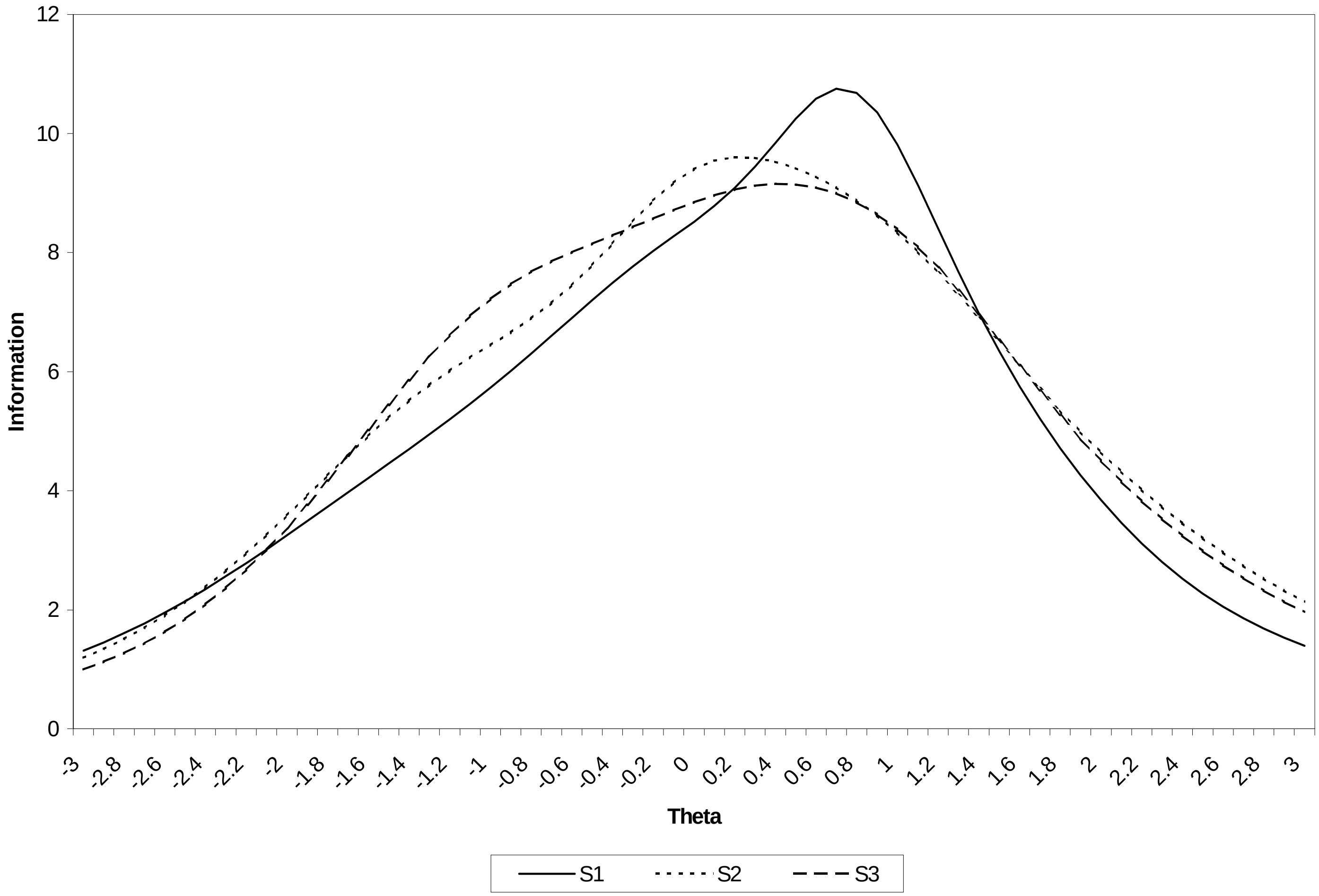
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure F.3— Test information curves for grade-8 geometry scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

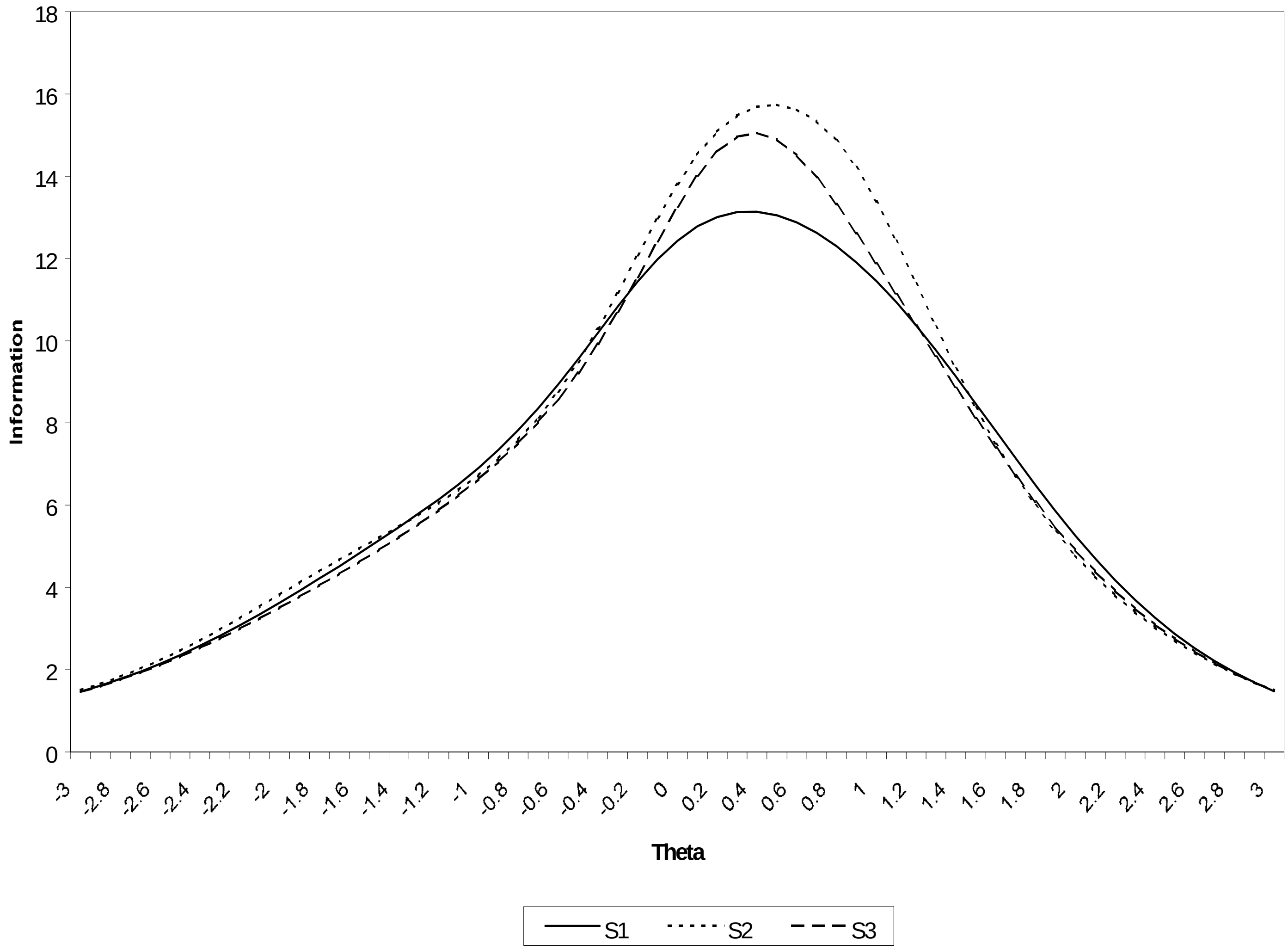
Figure F.4— Test information curves for grade-8 data analysis scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

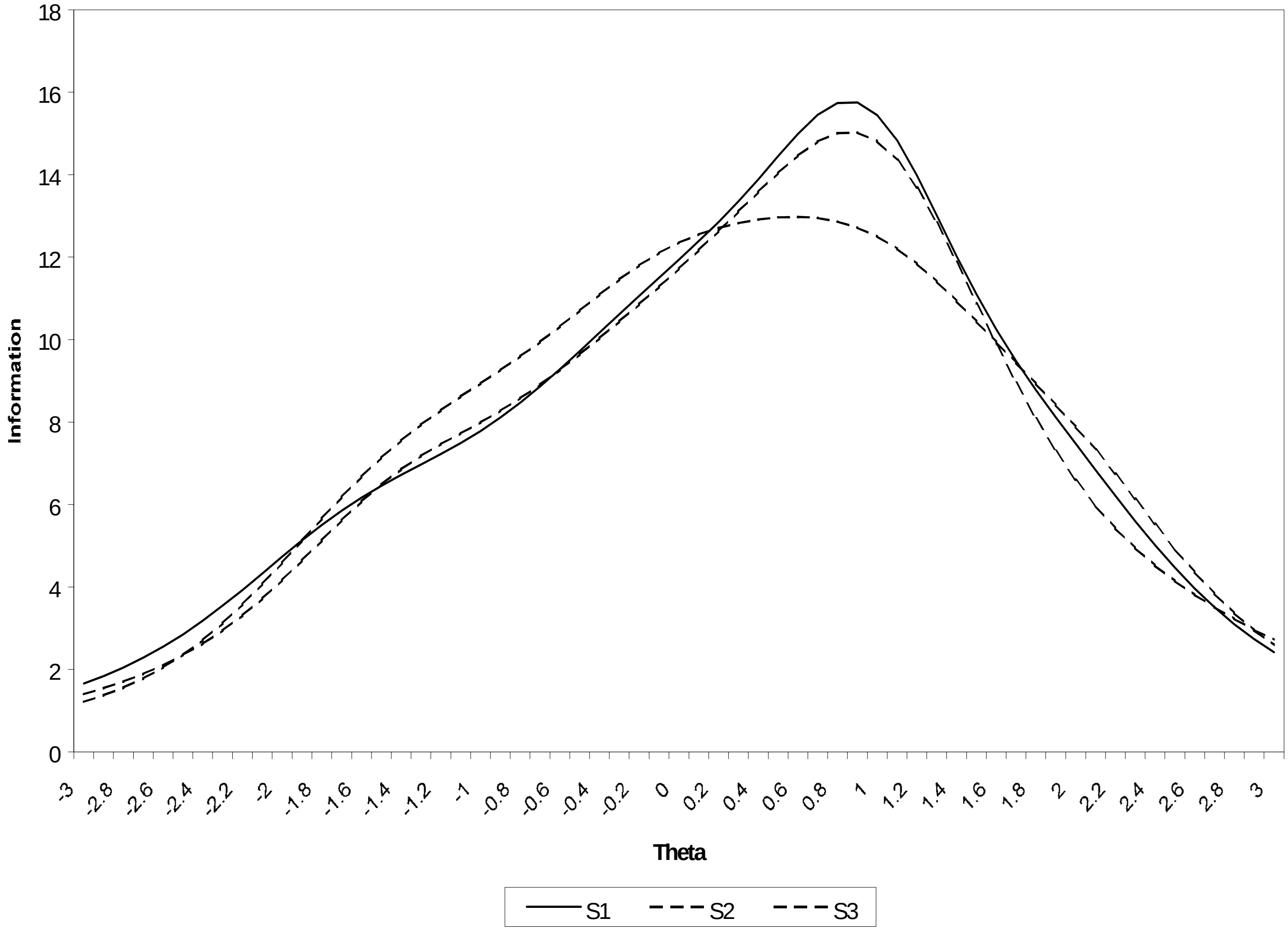
THE NATION'S REPORT CARD naep

Figure F.5 — Test information curves for grade-8 algebra and functions scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

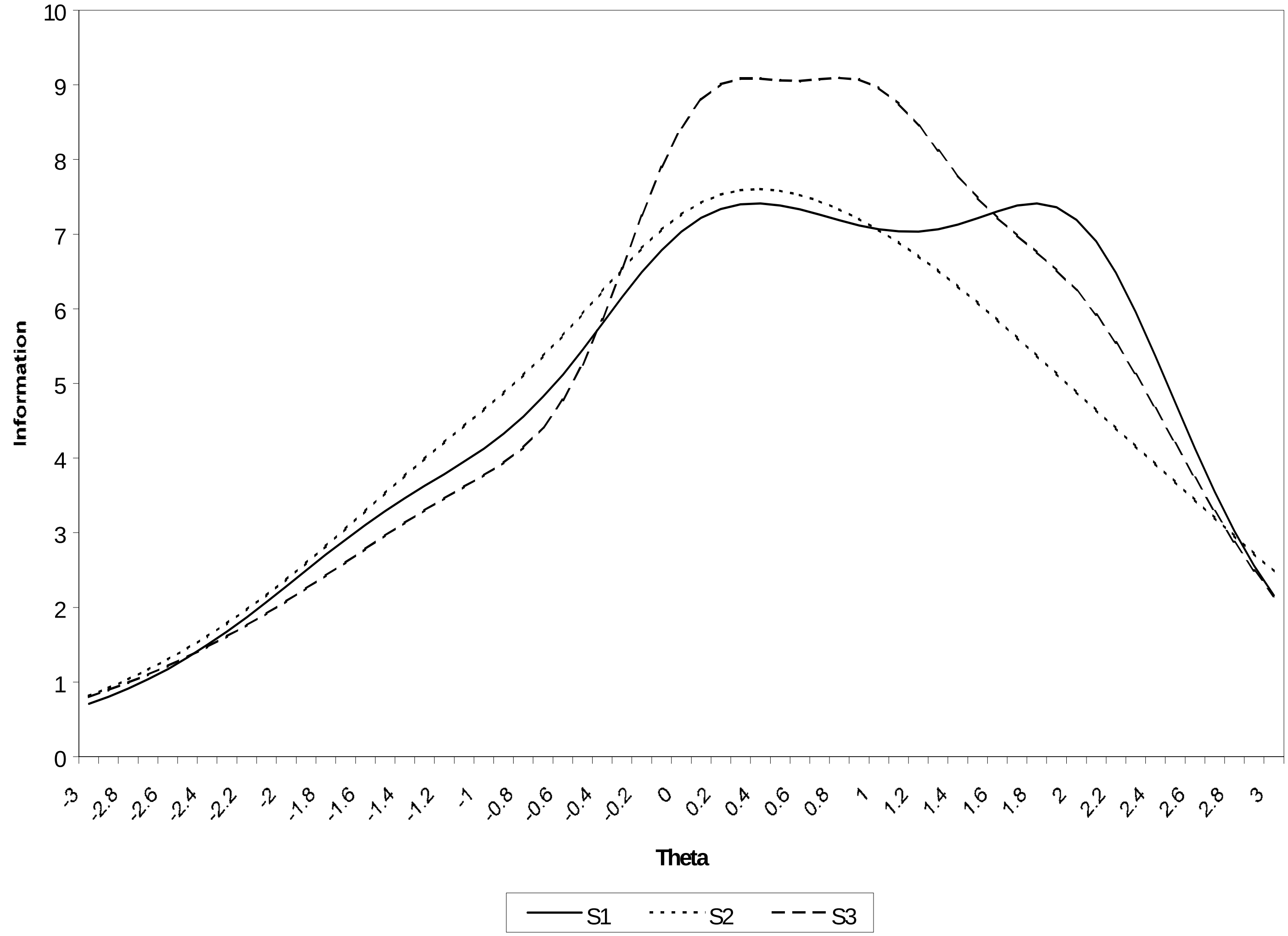
Figure F.6 — Test information curves for grade-12 numbers and operations scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

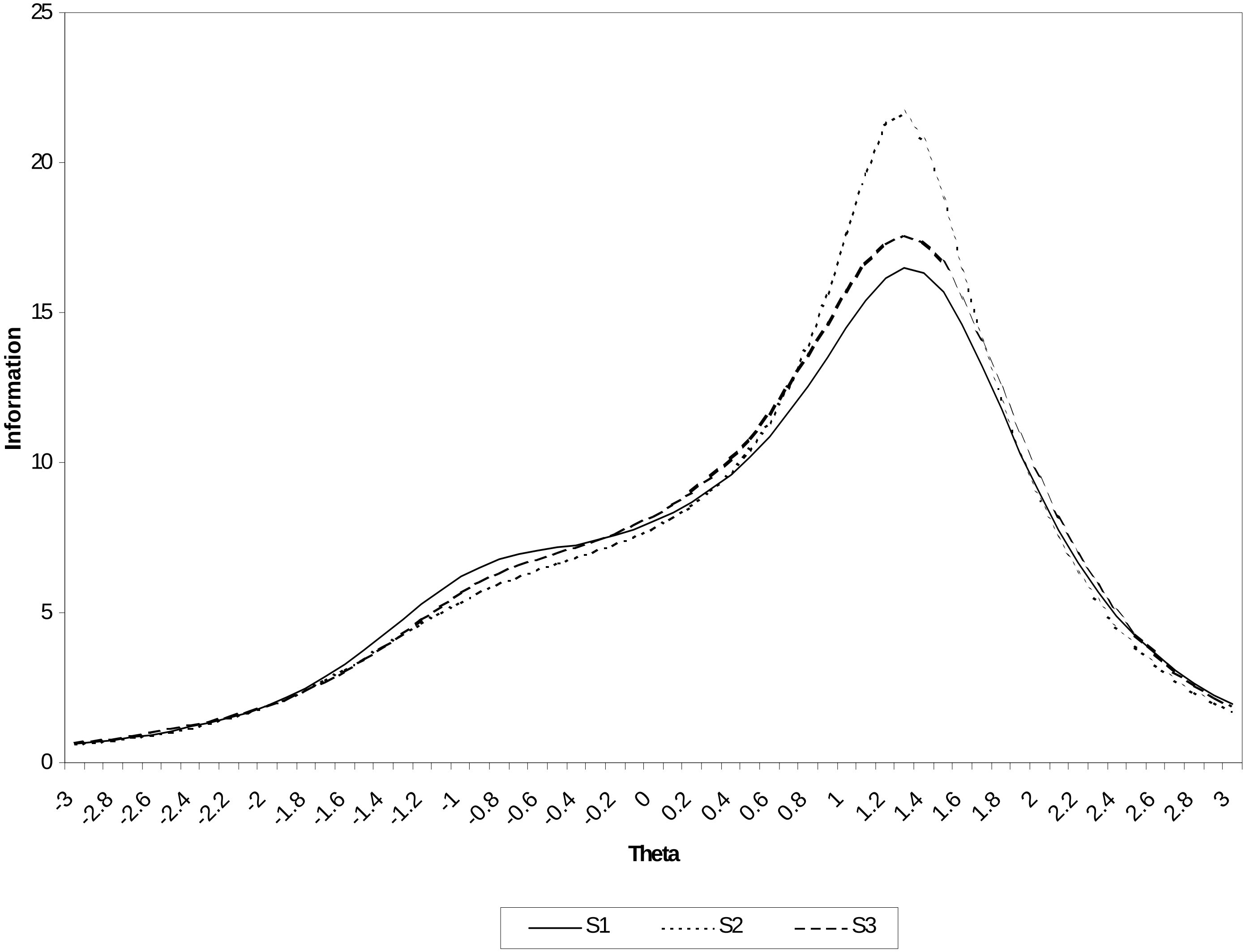
THE NATION'S REPORT CARD naep

Figure F.7 — Test information curves for grade-12 measurement scale



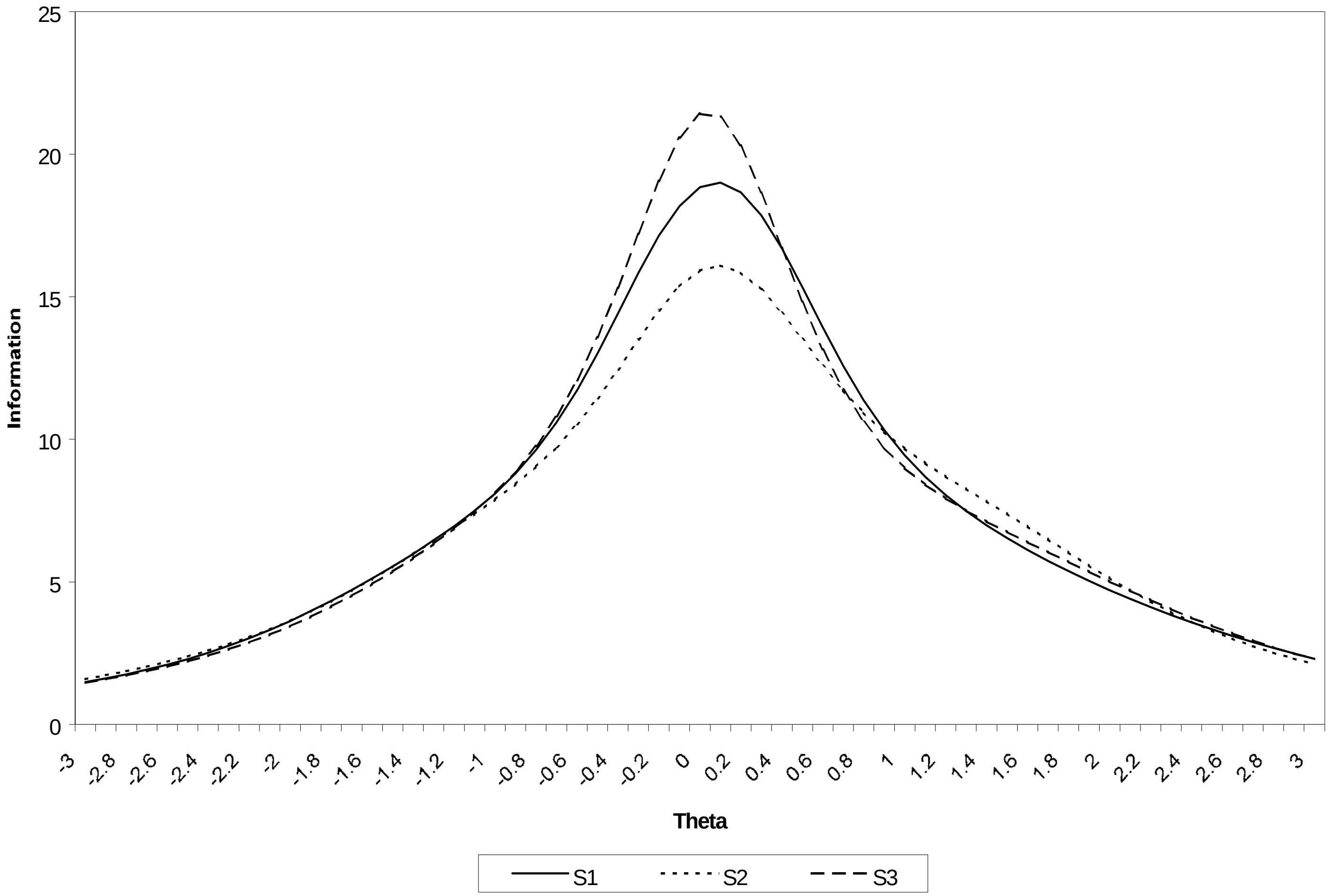
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure F.8 — Test information curves for grade-12 geometry scale



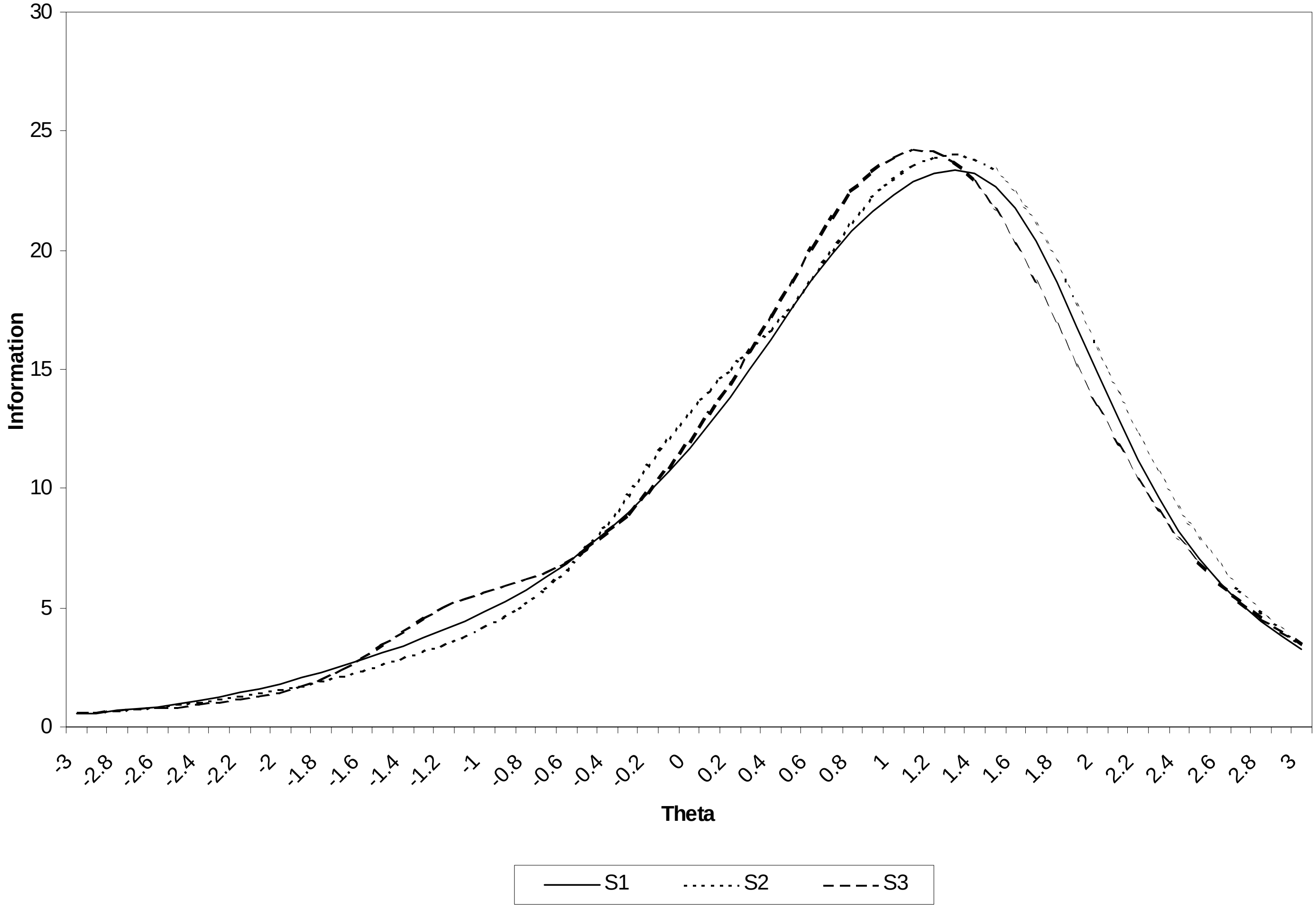
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure F.9 — Test information curves for grade-12 data analysis scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure F.10 — Test information curves for grade-12 algebra and functions scale



SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Appendix G

Standard Error Tables

Table G2.1 — Standard errors for percentage distribution of job titles of students with disabilities questionnaire respondents by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

	Grade 4	Grade 8	Grade 12
Principal/assistant principal	0.4	1.0	1.0
Special education teacher	3.1	3.1	3.0
Bilingual education/ESL teacher	0.4	0.1	0.1
Classroom teacher	3.0	1.4	0.8
Other	0.8	2.5	2.9
Blank	2.3	2.1	2.2

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G2.2 — Standard errors for percentage of students with disabilities by selected disability type by grade and gender: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Which of the following describes this student's disability?	Grade 4			Grade 8			Grade 12		
	All students	Male	Female	All students	Male	Female	All students	Male	Female
Learning disability	2.0	2.5	3.6	1.9	2.0	3.1	3.1	3.0	4.4
Speech/language impairment	1.6	2.0	2.5	1.2	1.3	1.8	1.3	1.2	2.7
Mental or cognitive impairment	1.8	2.1	3.4	2.0	1.9	3.2	2.6	2.5	4.4
Emotional disturbance	1.2	1.6	1.5	1.6	1.8	2.2	1.2	1.3	2.9
Hard of hearing	0.3	0.3	0.8	0.3	0.2	0.6	0.6	0.6	1.5
Deaf	0.2	—	0.6	0.2	0.3	0.2	0.5	0.6	0.8
Visual impairment/blindness	0.5	0.6	0.8	0.3	0.2	0.7	0.7	0.6	2.1
Orthopedic impairment	0.6	0.6	1.1	0.5	0.5	0.8	0.6	0.6	1.3
Autism	0.5	0.3	1.6	0.2	0.3	—	0.1	0.2	0.2
Traumatic brain injury	0.2	0.1	0.4	0.1	0.2	0.2	0.4	0.5	0.6
Other	0.9	1.1	1.1	0.7	1.0	1.4	1.0	1.2	1.4

— Not applicable because accommodations were not offered

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G2.3 — Standard errors for percentage distribution of degrees of disabilities by type of disability and grade: NAEP 1996 mathematics sample

	Grade 4			Grade 8			Grade 12		
What is the degree of this student's disability?	All students	Students with mental impairment	Students with other disabilities	All students	Students with mental impairment	Students with other disabilities	All students	Students with mental impairment	Students with other disabilities
Profound	0.3	1.5	0.2	0.5	2.4	0.3	1.0	2.8	0.7
Severe	1.9	4.5	2.0	1.4	5.3	1.6	1.8	3.2	1.8
Moderate	2.6	5.1	2.8	1.9	4.5	2.2	2.3	4.7	2.5
Mild	2.9	4.5	3.2	2.4	4.5	2.8	2.9	5.1	3.0

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G2.4 — Standard errors for percentage of time students with disabilities are mainstreamed in academic subjects by type of disability and grade: NAEP 1996 mathematics sample

What percentage of time is this student mainstreamed (i.e., with his/her nondisabled peers) in academic subjects (e.g., mathematics, reading/language arts, science)?	**Less than 40%**	**40% to 79%**	**At least 80%**
Grade 4			
All Students	2.4	2.0	2.8
Students with a:			
Learning disability	2.4	2.4	3.2
Cognitive impairment	6.6	4.5	5.1
Language impairment	4.4	3.6	5.2
Emotional disturbance	6.3	5.1	6.2
Grade 8			
All students	2.1	2.1	2.2
Students with a:			
Learning disability	2.3	2.3	2.2
Cognitive impairment	5.3	4.6	3.8
Language impairment	6.2	6.3	5.4
Emotional disturbance	5.7	4.9	5.1
Grade 12			
All students	2.4	2.1	2.6
Students with a:			
Learning disability	3.0	2.9	3.2
Cognitive impairment	4.3	4.2	2.6
Language impairment	8.3	6.7	5.8
Emotional disturbance	8.0	5.8	7.8

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G2.5 — Standard errors for percentage distribution of grade level of instruction in reading/language arts, mathematics, and science for students with disabilities by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

What grade level of instruction is the student currently receiving in:	Grade 4	Grade 8	Grade 12
Reading/language arts			
At/above grade level	2.6	2.6	2.4
One year below grade level	1.5	1.2	0.7
Two or more years below grade level	2.5	2.5	2.5
Missing	0.9	1.5	1.7
Mathematics			
At/above grade level	2.8	2.5	2.2
One year below grade level	2.2	1.3	0.8
Two or more years below grade level	1.5	2.7	2.6
Missing	1.0	1.5	2.3
Science			
At/above grade level	2.5	2.4	2.3
One year below grade level	1.4	0.7	0.9
Two or more years below grade level	1.7	2.0	3.4
Missing	1.4	1.8	3.9

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G2.6 — Standard errors for percentage of students with disabilities receiving the same curriculum content as nondisabled students at the same grade level, by grade level of instruction and grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Instructional Areas	Grade 4	Grade 8	Grade 12
Reading/language arts			
All students	2.9	2.5	2.3
Students receiving instruction:			
At/above grade level	1.3	1.5	1.3
Below grade level	3.4	3.8	3.6
Mathematics			
All students	2.2	2.3	2.8
Students receiving instruction:			
At/above grade level	0.7	1.2	1.1
Below grade level	3.9	3.4	3.8
Science			
All students	1.9	1.9	2.9
Students receiving instruction:			
At/above grade level	1.1	0.7	1.7
Below grade level	5.9	4.6	3.2

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G2.7 — Standard errors for percentage of students with disabilities receiving instruction in selected areas as part of their special education programs by grade: NAEP 1996 mathematics sample

Instructional Areas	Grade 4	Grade 8	Grade 12
Language development	2.8	2.5	3.3
Reading	2.4	2.7	3.2
Mathematics	2.9	2.4	2.6
Speech	1.7	1.2	1.1
Self-control and deportment	1.6	1.4	1.6
Personal care and basic life skills	1.6	1.5	1.9
Vocation education	0.6	1.5	2.6
Other	1.5	1.8	2.1

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G2.8 — Standard errors for percentage distribution of estimated grade level of performance in reading/language arts, mathematics, and science by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

At what grade level is this student currently performing in :	Grade 4	Grade 8	Grade 12
Reading/language arts			
Above grade level	0.4	0.3	0.5
At grade level	1.9	1.9	2.2
One year below grade level	1.8	1.9	1.3
Two or more years below grade level	2.4	2.6	2.6
I don't know.	1.2	1.6	1.7
Mathematics			
Above grade level	0.2	0.4	1.3
At grade level	2.4	2.3	1.8
One year below grade level	2.1	2.5	1.2
Two or more years below grade level	1.8	2.5	2.7
I don't know.	1.3	1.5	2.2
Science			
Above grade level	0.3	0.2	0.8
At grade level	2.9	2.6	1.9
One year below grade level	2.0	2.6	1.2
Two or more years below grade level	2.0	2.0	3.6
I don't know.	2.1	2.1	3.8

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G2.9 — Standard errors for percentage of students with disabilities using one or more accommodations for achievement testing by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Percentage of students with disabilities receiving:	Grade 4	Grade 8	Grade 12
Any accommodation	2.9	2.6	2.9
Presentation accommodation	2.8	2.9	3.0
Timing accommodation	3.0	2.5	2.8
Setting accommodation	2.8	2.6	3.0
Response accommodation	2.1	2.1	2.0

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G2.10 — Standard errors for percentage of students with disabilities receiving selected presentation accommodations and adaptations in achievement testing by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Type of presentation accommodation:	Grade 4	Grade 8	Grade 12
Read directions aloud	2.9	2.4	3.0
Read problems aloud	2.6	2.5	2.9
Assistance with directions	2.0	2.3	2.1
Use of taped version of test	0.9	0.5	1.5
Accommodation not listed on the questionnaire*	0.9	0.5	1.1

* At all three grades, less than half of 1 percent indicated each of the following accommodations: signing of directions; Braille edition of test; large-print edition of test; use of magnifying equipment.

NOTE: Questionnaire respondents were instructed to choose all responses that apply.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G2.11 — Standard errors for percentage of students with disabilities receiving selected types of timing accommodations in achievement testing by grade: NAEP 1996 mathematics sample

Type of timing accommodation	Grade 4	Grade 8	Grade 12
Extended time	3.0	2.6	2.8
More breaks during test	1.6	1.5	1.9
Test sessions over several days	1.3	1.4	1.8
Accommodation not listed on the questionnaire	0.9	0.4	0.2

NOTE: Questionnaire respondents were instructed to choose all responses that apply.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G2.12 — Standard errors for percentage of students with disabilities receiving selected setting accommodations in achievement testing by grade: NAEP 1996 mathematics assessment

Type of setting accommodation	Grade 4	Grade 8	Grade 12
Test in small group	2.7	2.8	3.2
Test individually	2.0	1.4	2.2
Accommodation not listed on the questionnaire	1.0	0.8	0.7

NOTE: Questionnaire respondents were instructed to choose all responses that apply.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G2.13 — Standard errors for percentage of students with disabilities receiving selected response accommodations used in achievement testing by grade: NAEP 1996 mathematics sample

Types of response accommodation	Grade 4	Grade 8	Grade 12
Oral responses	1.9	1.6	1.5
Use of braille/talking calculator	0.6	1.6	1.4
Pointing to answers	1.4	0.6	0.9
Use of computer to respond	0.8	1.0	1.5
Tape recording of answers	0.8	0.1	0.4
Use of typewriter	0.4	0	0.3
Accommodation not listed on questionnaire*	1.0	0.6	1.1

* At all three grades, less than half of 1 percent indicated each of the following accommodations: responding in Braille; responding in sign language; use of template; use of a special writing tool.

NOTE: Questionnaire respondents were instructed to choose all responses that apply.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G3.1 — Standard errors for percentage of students with disabilities in the national population included in the NAEP assessment, by grade and sample type: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

	Student Participation in NAEP				
		Assessed			
Sample type	N	% assessed without accommodations	% assessed with accommodations	Total % assessed	**% excluded**
Grade 4					
S1[1]	359	5.4	—	5.4	5.4
S2[2]	411	4.0	—	4.0	4.0
S3[3]	424	3.9	5.3	4.3	4.3
Grade 8					
S1[1]	310	4.0	—	4.0	4.0
S2[2]	524	4.4	—	4.4	4.4
S3[3]	557	4.2	3.5	3.4	3.4
Grade 12					
S1[1]	211	5.1	—	5.1	5.1
S2[2]	411	4.1	—	4.1	4.1
S3[3]	386	5.0	3.8	5.1	5.1

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

—Not applicable because accommodations were not offered

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G3.2 — Standard errors for percentage of students with disabilities who could meaningfully participate in NAEP mathematics without accommodations or adaptations by grade: NAEP 1996 mathematics sample

Could this student meaningfully participate in NAEP without accommodations?	Yes	No
Grade 4	2.3	2.3
Grade 8	2.7	2.7
Grade 12	2.7	2.7

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G3.3 — Standard errors for percentage distribution of participation status for students with disabilities by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

If accommodations or adaptations were available, how would this student participate in NAEP?	Without accommodations	With accommodations	Would not participate
Grade 4	2.3	3.0	2.7
Grade 8	2.5	2.4	2.1
Grade 12	2.4	2.6	3.0

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G3.4 — Standard errors for percentage of students with disabilities assessed in NAEP with each offered accommodation type by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Accommodation/adaptation type	Grade 4	Grade 8	Grade 12
Small group session	4.2	2.6	1.8
Extended time (regular session)	3.0	2.4	1.4
One-on-one testing	2.0	2.9	1.3
Directions read aloud (regular session)	1.5	0.6	2.6
Bilingual test booklet	0.2	0.0	0.0
Bilingual dictionary	0.0	0.0	0.0
Braille, large type booklet	0.0	0.5	0.0
Other accommodations	0.1	0.6	0.4

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G3.5 — Standard errors for percentage distribution of reasons for exclusion of students with disabilities from NAEP assessment, by sample type and grade: NAEP 1996 mathematics sample

Sample Type	N	Reason for Exclusion: Stated in IEP and judged to be impaired	Stated in IEP (only)	Judged to be impaired (only)	Neither reason
Grade 4					
S1[1]	143	6.8	5.7	2.4	6.9
S2[2]	190	7.0	4.6	0.9	5.3
S3[3]	80	8.1	6.1	2.1	4.3
Grade 8					
S1[1]	112	6.6	5.8	3.0	5.1
S2[2]	161	5.7	2.4	2.3	5.5
S3[3]	152	7.8	10.3	4.4	6.2
Grade 12					
S1[1]	89	7.7	4.5	3.4	5.9
S2[2]	199	5.3	3.8	1.4	4.2
S3[3]	146	6.0	5.0	1.4	4.4

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G3.6 — Standard errors for percentage of students included in NAEP by degree of disability, by grade and sample type: NAEP 1996 mathematics sample

What is the degree of this student's disability?	Mild			Moderate			Severe/Profound		
	N	% assessed	% excluded	N	% assessed	% excluded	N	% assessed	% excluded
Grade 4									
S1[1]	151	4.7	4.7	112	7.5	7.5	48	12.1	12.1
S2[2]	153	7.9	7.9	136	6.4	6.4	52	9.3	9.3
S3[3]	160	3.5	3.5	117	8.5	8.5	22	12.4	12.4
Grade 8									
S1[1]	117	5.5	5.5	78	6.0	6.0	38	3.7	3.7
S2[2]	199	4.2	4.2	122	7.3	7.3	58	6.9	6.9
S3[3]	207	2.8	2.8	173	4.7	4.7	58	10.6	10.6
Grade 12									
S1[1]	78	6.1	6.1	63	11.1	11.1	25	26.4	26.4
S2[2]	159	5.7	5.7	109	7.3	7.3	60	6.4	6.4
S3[3]	130	5.8	5.8	99	8.2	8.2	36	9.0	9.0

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G3.7 — Standard errors for percentage of students with disabilities included in NAEP by percentage of time mainstreamed, by sample type and grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

What percentage of time is this student mainstreamed in academic subjects?	Less than 50%			50% or More		
	N	% assessed	% excluded	N	% assessed	% excluded
Grade 4						
S1[1]	68	5.7	5.7	243	5.2	5.2
S2[2]	90	5.6	5.6	253	5.1	5.1
S3[3]	67	9.2	9.2	235	4.1	4.1
Grade 8						
S1[1]	76	6.9	6.9	164	4.6	4.6
S2[2]	154	9.3	9.3	233	3.6	3.6
S3[3]	173	5.2	5.2	275	2.7	2.7
Grade12						
S1[1]	46	5.2	5.2	126	7.3	7.3
S2[2]	137	7.3	7.3	193	5.1	5.1
S3[3]	111	10.4	10.4	156	4.8	4.8

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G3.8 — Standard errors for percentage of students with disabilities included in NAEP by grade level of instruction and curriculum content in mathematics, by sample type and grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

	At/above grade level			Below grade level					
	All students			Students with same curriculum as nondisabled students			Students with different curriculum from non-disabled students		
Sample type	N	% assessed	% excluded	N	% assessed	% excluded	N	% assessed	% excluded
Grade 4									
S1[1]	175	6.1	6.1	53	10.5	10.5	77	8.0	8.0
S2[2]	169	5.8	5.8	81	7.2	7.2	72	4.4	4.4
S3[3]	168	3.7	3.7	52	7.7	7.7	74	8.0	8.0
Grade 8									
S1[1]	104	4.0	4.0	50	8.6	8.6	83	7.0	7.0
S2[2]	180	4.2	4.2	61	7.3	7.3	130	5.8	5.8
S3[3]	250	3.6	3.6	72	7.7	7.7	118	7.1	7.1
Grade 12									
S1[1]	43	4.9	4.9	50	9.9	9.9	51	4.0	4.0
S2[2]	78	5.1	5.1	46	9.8	9.8	127	4.0	4.0
S3[3]	58	7.1	7.1	52	7.2	7.2	106	11.7	11.7

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G3.9 — Standard errors for percentage of students with disabilities assessed in 1996 mathematics by accommodation/adaptation status, by sample type and grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD

Are accommodations/ adaptations used for achievement testing for this student?	Yes				No				IED says student cannot be tested			
	Participation in NAEP				Participation in NAEP				Participation in NAEP			
	N	% assessed standard	% assessed with accom.	% excluded	N	% assessed standard	% assessed with accom.	% excluded	N	% assessed standard	% assessed with accom.	% excluded
Grade 4												
S1[1]	97	9.0	—	9.0	131	5.7	—	5.7	80	4.7	—	4.7
S2[2]	112	8.1	—	8.1	136	7.1	—	7.1	89	3.3	—	3.3
S3[3]	143	5.3	6.5	5.3	101	3.8	4.2	3.8	56	9.9	10.0	10.0
Grade 8												
S1[1]	75	7.2	—	7.2	122	5.4	—	5.4	40	4.1	—	4.1
S2[2]	158	6.5	—	6.5	153	3.6	—	3.6	71	9.2	—	9.2
S3[3]	183	4.9	7.3	4.9	181	3.0	4.3	3.0	81	5.5	2.0	5.5
Grade 12												
S1[1]	51	11.4	—	11.4	73	7.7	—	7.7	45	2.5	—	2.5
S2[2]	135	6.1	—	6.1	88	6.8	—	6.8	99	5.3	—	5.3
S3[3]	94	6.0	11.0	6.0	86	6.6	3.2	6.6	85	1.3	0.6	1.3

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

— Not applicable because accommodations were not offered

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G4.1 — Standard errors for percentage distribution of job titles of limited English proficient student questionnaire respondents by grade: NAEP 1996 mathematics sample

Who filled out the questionnaire?	Grade 4	Grade 8	Grade 12
Bilingual education teacher	3.6	4.7	4.1
Classroom teacher	5.3	3.5	2.8
Special education teacher	1.0	0.7	1.2
Principal	0.2	3.8	1.6
Other	1.4	2.4	7.0
Missing	5.0	6.1	5.4

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G4.2 — Standard errors for percentage distribution of time living in U.S. for limited English proficient students by grade: NAEP 1996 mathematics sample

How long has this student lived in the U.S.?	Grade 4	Grade 8	Grade 12
All his/her life	3.8	2.2	2.1
More than 5 years	1.8	2.9	3.1
3-5 years	2.6	2.8	2.4
Less than 3 years	2.0	3.7	3.2
I don't know.	2.1	4.7	3.3

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G4.3 — Standard errors for percentage distribution of first or native language for limited English proficient students by grade: NAEP 1996 mathematics sample

What is student's first or native language?	Grade 4	Grade 8	Grade 12
Spanish	3.2	4.5	4.7
Another language	3.2	4.5	4.7

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G4.4 — Standard errors for percentage distribution of regularity of school attendance in U.S. or another country for limited English proficient students by grade: NAEP 1996 mathematics sample

Since reaching school age, how regularly has student attended school in U.S. or another country?	Grade 4	Grade 8	Grade 12
Continuously	2.7	4.6	7.4
Intermittently	0.9	1.3	2.2
Little or not at all	0.9	0.7	1.1
I don't know.	1.8	4.9	7.4

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G4.5 — Standard errors for percentage distribution of number of years limited English proficient students have been enrolled in a school where English is primary language of instruction by grade: NAEP 1996 mathematics sample

Counting this year, how many years has this student been enrolled in a school where English is the primary language of instruction?	Grade 4	Grade 8	Grade 12
1 years	1.9	2.2	1.8
2 years	2.4	2.7	2.8
3-5 years	3.8	3.4	3.2
More than 5 years	0.9	3.6	3.8
English not school's primary language	4.2	0.6	0.4
I don't know.	2.1	4.1	3.5

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G4.6 — Standard errors for percentage distribution of years of enrollment in academic instruction specially designed for limited English proficient students by grade: NAEP 1996 mathematics sample

Counting this year, how many years has this student been receiving academic instruction specially designed for students with limited English proficiency?	Grade 4	Grade 8	Grade 12
1 year	1.7	2.8	2.5
2 years	1.8	3.2	2.6
3-5 years	3.4	3.4	3.4
More than 5 years	1.0	3.4	1.7
Not receiving special instruction	2.8	3.1	4.2

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G4.7 — Standard errors for percentage distribution of years of academic instruction in English for limited English proficient students by grade: NAEP 1996 mathematics sample

Counting this year, how many years has this student been receiving academic instruction primarily in English?	Grade 4	Grade 8	Grade 12
1 year	2.3	2.4	1.8
2 years	2.3	2.3	2.8
3-5 years	3.9	2.8	3.0
More than 5 years	0.8	3.2	3.4
Not receiving instruction in English	3.5	2.2	0.8
I don't know.	1.9	4.0	3.0

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G4.8 — Standard errors for percentage of limited English proficient students receiving specially designed instruction in selected content areas by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD

Instructional Area	Grade 4	Grade 8	Grade 12
Reading/language arts			
Special instruction in English	3.0	5.0	6.5
Native language instruction	2.7	2.9	—
No special instruction	3.2	3.5	6.5
Mathematics			
Special instruction in English	3.6	4.2	4.6
Native language instruction	3.5	1.6	4.2
No special instruction	3.6	4.0	5.5
Science			
Special instruction in English	3.2	5.0	4.6
Native language instruction	3.8	1.5	3.7
No special instruction	4.5	4.5	5.3

— Not applicable because accommodations were not offered

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G4.9 — Standard errors for percentage distribution of limited English proficient students by grade level of English language instruction in reading/language arts, mathematics, and science by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

What grade level of instruction in the English language is this student currently receiving in:	Grade 4	Grade 8	Grade 12
Reading/language arts			
Above grade level	0.0	0.0	0.0
At grade level	4.3	4.3	4.0
One year below grade level	3.2	1.6	2.2
Two or more years below grade level	2.7	3.9	4.6
Mathematics			
Above grade level	0.4	0.4	0.0
At grade level	3.0	5.2	3.9
One year below grade level	2.2	2.8	3.2
Two or more years below grade level	1.8	4.0	5.3
Science			
Above grade level	0.1	0.0	0.0
At grade level	2.5	5.3	5.1
One year below grade level	2.3	1.6	2.4
Two or more years below grade level	1.8	4.6	4.7

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G4.10 — Standard errors for percentage distribution of estimated grade level of performance in reading/language arts, mathematics, and science for limited English proficient students by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

At what grade level is this student currently performing in the English language in:	Grade 4	Grade 8	Grade 12
Reading/language arts			
Above grade level	0.4	0.6	1.5
At grade level	3.8	3.4	3.4
One year below grade level	2.3	2.3	2.3
Two or more years below grade level	3.8	3.6	4.5
I don't know.	0.5	4.0	2.8
Mathematics			
Above grade level	0.8	0.8	2.2
At grade level	4.0	3.6	4.2
One year below grade level	3.1	2.6	2.6
Two or more years below grade level	2.2	3.8	4.0
I don't know.	1.1	3.2	2.6
Science			
Above grade level	0.4	0.5	1.9
At grade level	3.6	4.0	4.6
One year below grade level	2.3	2.4	2.5
Two or more years below grade level	2.9	4.0	4.1
I don't know.	0.9	4.0	3.3

NOTE: Because of rounding, percentages may not add to 100.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G4.11 — Standard errors for percentage distribution for limited English proficient students of whether accommodations or adaptations are used for achievement testing by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Are any accommodations or adaptations used for achievement testing for this student?	Grade 4	Grade 8	Grade 12
Yes	3.8	3.5	3.3
No	3.8	3.5	3.3

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G4.12 — Standard errors for percentage of limited English proficient students receiving selected accommodations/adaptations in achievement testing by grade: NAEP 1996 mathematics sample

THE NATION'S REPORT CARD naep

Accommodation/adaptation	Grade 4	Grade 8	Grade 12
Native language version of test	3.0	2.2	1.6
English/native language dictionary	1.4	1.5	2.9
Word lists or glossaries	2.1	0.7	2.3
Extended time	2.0	2.9	2.4
Help with directories and questions	1.7	2.4	1.9
Directions read aloud twice in English	3.0	2.6	1.4
Questions read aloud twice in English	2.3	1.9	1.6
Other	0.9	2.1	2.0

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G5.1 — Standard errors for percentage of limited English proficient students included in NAEP, by sample type and grade: 1996 NAEP mathematics sample

Sample Type	Assessed			Excluded
	Without accommodations	With accommodations	Total	**Total**
Grade 4				
S1[1]	8.2	—	8.2	8.2
S2[2]	5.3	—	5.3	5.3
S3[3]	6.7	7.2	7.0	7.0
Grade 8				
S1[1]	9.6	—	9.6	9.6
S2[2]	5.4	—	5.4	5.4
S3[3]	4.0	4.6	4.4	4.4
Grade 12				
S1[1]	6.5	—	6.5	6.5
S2[2]	6.8	—	6.8	6.8
S3[3]	4.0	2.0	3.4	3.4

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

— Not applicable because of accommodations were not offered.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G5.2 — Standard errors for percentage of limited English proficient students who could meaningfully participate in NAEP without accommodations or adaptations by grade: 1996 NAEP mathematics sample

In your judgement, could this student meaningfully participate in NAEP with accommodations?	Grade 4		Grade 8		Grade 12	
	Yes	No	Yes	No	Yes	No
Mathematics	3.6	3.6	3.1	3.1	6.1	6.1

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G5.3 — Standard errors for percentage distribution of limited English proficient students who would participate in NAEP, by accommodation status and grade: 1996 NAEP mathematics sample

How would this student participate in NAEP?	Grade 4	Grade 8	Grade 12
In English			
Without accommodation/adaptation	3.2	4.0	5.7
With accommodation/adaptation	2.7	2.8	6.0
Total	3.6	3.2	3.0
In Native Language			
Without accommodation/adaptation	4.8	2.4	2.1
With accommodation/adaptation	3.5	3.5	0.2
Total	4.8	4.1	2.2
Student Would Not Participate	2.8	2.6	1.7

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G5.4 — Standard errors for percentage of limited English proficient students assessed with each of the offered accommodations/adaptations by grade: 1996 NAEP mathematics sample

Accommodation/adaptation	Grade 4	Grade 8	Grade 12
Bilingual test booklet	7.3	4.8	*
One-on-one testing	0.7	0.4	1.3
Extended time (regular session)	1.0	0.4	0.9
Small group session	1.0	—[1]	0.4
Directions read aloud Regular session	0.4	1.7	0.5
Bilingual dictionary	—[1]	0.7	—[1]
Other	—	—[1]	0.7

—[1] Estimated standard errors not defined since estimated standard errors for percentage is exactly zero

* Bilingual test booklet was not offered at grade 12.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G5.5 — Standard errors for percentages of limited English proficient students assessed by years enrolled in English-language school, by sample type and grade: 1996 NAEP mathematics sample

Counting this year, how many years has this student been enrolled in a school where English is the primary language of instruction?	Less than 2 years			2 years or more		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4						
S1[1]	26	3.8	3.8	82	8.2	8.2
S2[2]	63	11.4	11.4	161	5.9	5.9
S3[3]	35	7.1	7.1	170	9.4	9.4
Grade 8						
S1[1]	15	——	——	51	9.3	9.3
S2[2]	26	7.9	7.9	150	7.0	7.0
S3[3]	24	16.4	16.4	73	7.0	7.0
Grade 12						
S1[1]	1	——	——	30	9.0	9.0
S2[2]	11	——	——	147	9.6	9.6
S3[3]	10	——	——	107	4.2	4.2

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

—— Indicates insufficient sample size to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G5.6 — Standard errors for percentages of limited English proficient students assessed by years receiving academic instruction in English, by sample type and grade: 1996 NAEP mathematics sample

Counting this year, how many years has this student been receiving academic instruction primarily in English?	Less than 3 years			3 years or more		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4						
S1[1]	66	11.0	11.0	36	5.4	5.4
S2[2]	124	6.0	6.0	107	5.4	5.4
S3[3]	96	12.6	12.6	108	6.5	6.5
Grade 8						
S1[1]	35	18.0	18.0	26	5.9	5.9
S2[2]	71	7.2	7.2	100	5.7	5.7
S3[3]	58	8.2	8.2	51	8.6	8.6
Grade 12						
S1[1]	5	——	——	26	7.4	7.4
S2[2]	46	11.0	11.0	112	9.5	9.5
S3[3]	30	8.0	8.0	89	3.9	3.9

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

—— Indicates insufficient sample size to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G5.7 — Standard errors for percentages of limited English proficient students assessed by whether accommodations or adaptations are normally used, by sample type and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

Are any accommodation/ adaptations used for achievement testing for this student?	YES				NO			
	N	Assessed standard	Assessed with accommodations	Excluded	N	Assessed standard	Assessed with accommodations	Excluded
Grade 4								
S1[1]	31	11.0	——	11.0	95	10.9	——	10.9
S2[2]	81	6.9	——	6.9	169	8.7	——	8.7
S3[3]	109	7.2	8.3	11.5	103	6.5	4.0	5.0
Grade 8								
S1[1]	19	——	——	——	60	11.4	——	11.3
S2[2]	63	6.9	——	6.9	137	8.0	——	8.0
S3[3]	47	10.3	11.1	9.8	114	3.5	1.8	4.1
Grade 12								
S1[1]	2	——	——	——	31	9.4	——	9.4
S2[2]	30	7.0	——	7.0	138	10.4	——	10.4
S3[3]	30	9.1	8.5	9.7	102	3.7	2.6	2.8

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

—— Indicates insufficient sample size to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G5.8 — Standard errors for percentages of limited English proficient students assessed, by preferred language of assessment, sample type and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

If accommodations/adaptations were available, how would this student participate in NAEP?	In English			In native language		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4						
S1[1]	50	5.2	5.2	56	12.7	12.7
S2[2]	109	6.8	6.8	87	2.9	2.9
S3[3]	116	4.4	4.4	68	13.8	13.8
Grade 8						
S1[1]	56	6.3	6.3	10	——	——
S2[2]	109	3.6	3.6	36	11.7	11.7
S3[3]	105	5.1	5.1	44	5.6	5.6
Grade 12						
S1[1]	21	6.0	6.0	1	——	——
S2[2]	136	9.6	9.6	16	——	——
S3[3]	111	3.7	3.7	9	——	——

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

—— Indicates insufficient sample size to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G5.9 — Standard errors for percentages of limited English proficient students assessed, by native language, sample type and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

What is the student's first or native language?	Spanish			Another Language		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4						
S1[1]	107	12.8	12.8	15	——	——
S2[2]	178	6.8	6.8	59	7.0	7.0
S3[3]	159	8.7	8.9	52	7.4	7.4
Grade 8						
S1[1]	28	14.0	14.0	31	14.8	14.8
S2[2]	172	7.2	7.2	52	9.3	9.3
S3[3]	110	4.9	4.9	44	8.5	8.5
Grade 12						
S1[1]	16	——	——	14	——	——
S2[2]	108	10.0	10.0	48	10.6	10.6
S3[3]	46	6.4	6.4	81	4.2	4.2

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

—— Indicates insufficient sample size to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G5.10 — Standard errors for percentages of limited English proficient students assessed by type of mathematics instruction, sample type and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

Type of mathematics instruction during the current year?	No special instruction			Specially designed instruction in English			Native language instruction		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4									
S1[1]	34	9.6	9.6	34	7.5	7.5	33	44.6	44.6
S2[2]	95	6.6	6.6	79	9.1	9.1	73	2.3	2.3
S3[3]	91	6.5	6.5	62	4.4	4.4	56	14.3	14.3
Grade 8									
S1[1]	32	7.8	7.8	28	22.2	22.2	6	——	——
S2[2]	129	4.6	4.6	47	9.3	9.3	24	13.8	13.8
S3[3]	77	6.9	6.9	73	5.0	5.0	14	—	—
Grade 12									
S1[1]	32	8.3	8.3	0	——	——	0	——	——
S2[2]	91	5.9	5.9	53	16.2	16.2	19	——	——
S3[3]	94	4.3	4.3	24	10.5	10.5	5	——	——

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

—— Indicates insufficient sample size to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G5.11 — Standard errors for percentages of limited English proficient students assessed, by grade level of mathematics instruction in English, sample type and grade: 1996 NAEP mathematics sample

THE NATION'S REPORT CARD naep

	Grade level of instruction in mathematics								
	In English						In native language		
	At or above grade level			Below grade level			All students		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4									
S1[1]	63	5.2	5.2	23	15.7	15.7	37	26.6	26.6
S2[2]	152	6.0	6.0	23	15.6	15.6	68	2.8	2.8
S3[3]	155	8.2	8.2	25	13.6	13.6	25	16.8	16.8
Grade 8									
S1[1]	45	4.9	4.9	21	23.7	23.7	0	——	——
S2[2]	157	6.9	6.9	28	14.0	14.0	5	——	——
S3[3]	122	3.6	3.6	35	12.9	12.9	4	——	——
Grade 12									
S1[1]	21	8.0	8.0	10	——	——	0	——	——
S2[2]	99	13.0	13.0	50	9.9	9.9	4	——	——
S3[3]	71	5.1	5.1	40	7.0	7.0	1	——	——

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

—— Indicates insufficient sample size to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G5.12 — Standard errors for percentages of limited English proficient students assessed by estimated level of mathematics performance in English, by sample type and grade: 1996 NAEP mathematics sample

	Grade level of performance in mathematics								
	In English						In native language		
	At or above grade level			Below grade level			All students		
	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded	N	Percentage assessed	Percentage excluded
Grade 4									
S1[1]	35	7.1	7.1	47	6.6	6.6	31	42.1	42.1
S2[2]	114	6.2	6.2	59	7.4	7.4	65	0.6	0.6
S3[3]	101	5.2	5.2	74	7.1	7.1	30	14.7	14.7
Grade 8									
S1[1]	19	——	——	44	19.9	19.9	0	——	——
S2[2]	71	5.7	5.7	67	9.7	9.7	4	——	——
S3[3]	55	5.3	5.3	50	10.2	10.2	2	——	——
Grade 12									
S1[1]	16	——	——	8	——	——	0	——	——
S2[2]	64	9.6	9.6	60	11.9	11.9	1	——	——
S3[3]	46	5.3	5.3	34	11.4	11.4	0	——	——

[1] 1990-1996 operational inclusion criteria

[2] 1996 revised inclusion criteria

[3] Revised inclusion criteria plus accommodations in testing

—— Indicates insufficient sample size to permit a reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

THE NATION'S REPORT CARD naep

Table G7.1 — Standard errors for means with significant differences between samples S2 and S3: Science, grades 4 and 8

	Students from sample S2	Students from sample S3
LEP Students (some are also SD)		
Grade 4/physical science	4.6	3.3
Grade 8/composite	3.5	3.9
Grade 8/earth science	3.7	5.2
LEP Students Not also SD		
Grade 8/earth science	3.8	5.1
Combined SD and LEP Students		
Grade 8/earth science	2.5	3.4

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table G7.2 — Standard errors for grade-4 science means overall and by gender

	Students from sample S2	Students from sample S3
Composite	**0.8**	**1.2**
Female	**0.9**	**1.3**
Male	**0.9**	**1.3**
Physical science	1.0	1.3
Female	1.2	1.4
Male	1.1	1.5
Earth science	0.8	1.3
Female	1.0	1.5
Male	0.9	1.5
Life science	0.9	1.4
Female	1.0	1.6
Male	1.1	1.5

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table G7.3 — Standard errors for grade-8 science means overall and by gender

	Students from sample S2	Students from sample S3
Composite	**0.9**	**1.0**
Female	**1.1**	**1.0**
Male	**1.0**	**1.3**
Physical science	0.9	1.0
Female	1.1	1.2
Male	1.2	1.2
Earth science	1.0	1.1
Female	1.2	1.1
Male	1.2	1.6
Life science	1.1	1.1
Female	1.3	1.1
Male	1.2	1.5

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table G7.4 — Standard errors for grade-12 science means overall and by gender

	Students from sample S2	Students from sample S3
Composite	**0.9**	**0.9**
Female	**0.9**	**1.1**
Male	**1.2**	**1.1**
Physical science	1.0	0.9
Female	1.0	1.1
Male	1.4	1.1
Earth science	0.9	1.0
Female	1.0	1.3
Male	1.4	1.2
Life science	0.9	1.1
Female	1.2	1.3
Male	1.1	1.3

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table G7.5 — Standard errors for grade-4 science means by race/ethnicity

	Students from sample S2	Students from sample S3
Composite		
Black	**1.9**	**2.4**
American Indian	**3.8**	**5.9**
Asian/Pacific Islander	**3.6**	**4.6**
Hispanic	**1.7**	**2.6**
White	**0.9**	**1.2**
Physical science		
Black	2.2	1.9
American Indian	3.5	7.4
Asian/Pacific Islander	3.8	5.1
Hispanic	2.2	2.3
White	1.2	1.3
Earth science		
Black	2.2	3.7
American Indian	4.6	6.5
Asian/Pacific Islander	4.1	5.2
Hispanic	2.2	3.1
White	1.0	1.4
Life science		
Black	2.0	2.4
American Indian	4.0	4.7
Asian/Pacific Islander	4.0	5.1
Hispanic	2.0	3.3
White	1.0	1.4

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table G7.6 — Standard errors for grade-8 science means by race/ethnicity

	Students from sample S2	Students from sample S3
Composite		
Black	**1.1**	**2.0**
American Indian	**4.1**	**3.4**
Asian/Pacific Islander	**3.1**	**4.6**
Hispanic	**1.7**	**2.7**
White	**1.1**	**0.9**
Physical science		
Black	1.3	2.1
American Indian	4.9	3.7
Asian/Pacific Islander	3.4	4.4
Hispanic	2.0	2.7
White	1.2	1.1
Earth science		
Black	1.7	2.2
American Indian	4.5	3.6
Asian/Pacific Islander	3.5	4.5
Hispanic	1.7	2.8
White	1.2	1.2
Life science		
Black	1.3	2.5
American Indian	3.9	3.8
Asian/Pacific Islander	3.4	5.3
Hispanic	2.2	3.1
White	1.2	1.0

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table G7.7 — Standard errors for grade-12 science means by race/ethnicity

	Students from sample S2	Students from sample S3
Composite		
Black	**1.5**	**1.8**
American Indian	**4.7**	**——**
Asian/Pacific Islander	**2.9**	**4.5**
Hispanic	**2.3**	**2.4**
White	**1.0**	**1.0**
Physical science		
Black	1.8	2.3
American Indian	6.1	——
Asian/Pacific Islander	2.8	5.4
Hispanic	2.5	2.6
White	1.1	1.0
Earth science		
Black	1.9	2.0
American Indian	4.1	——
Asian/Pacific Islander	3.0	3.9
Hispanic	2.4	2.5
White	1.0	1.2
Life science		
Black	1.4	2.0
American Indian	5.2	——
Asian/Pacific Islander	3.5	4.9
Hispanic	2.7	2.8
White	1.1	1.1

—— Insufficient sample size to permit reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table G7.8 — Standard errors for mathematics significant differences

	Students from sample S1	Students from sample S2	Students from sample S3
All Students			
Grade 12/geometry	1.5	1.2	1.5
Female Students			
Grade 12/geometry	1.7	1.7	1.4
Asian/Pacific Islander Students			
Grade 12/algebra	6.3	4.0	4.1

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G7.9 — Standard errors for grade-4 mathematics means overall and by gender

	Students from sample S1	Students from sample S2	Students from sample S3
Composite	**1.4**	**1.2**	**1.1**
Female	**1.5**	**1.1**	**1.3**
Male	**1.6**	**1.6**	**1.1**
Numeric operations	1.6	1.4	1.2
Female	1.7	1.3	1.4
Male	1.7	1.8	1.2
Measurement	1.5	1.4	1.5
Female	1.7	1.4	1.9
Male	1.8	1.7	1.3
Geometry	1.3	1.4	1.1
Female	1.4	1.2	1.3
Male	1.5	1.8	1.2
Data analysis	1.7	1.6	1.4
Female	1.9	1.7	1.5
Male	1.9	2.0	1.8
Algebra	1.5	1.5	1.1
Female	1.7	1.3	1.5
Male	1.7	1.9	1.5

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G7.10 — Standard errors for grade-8 mathematics means overall and by gender

	Students from sample S1	Students from sample S2	Students from sample S3
Composite	**1.5**	**1.2**	**1.1**
Female	**1.5**	**1.4**	**1.1**
Male	**1.8**	**1.4**	**1.4**
Numeric operations	1.5	1.2	0.9
Female	1.5	1.5	1.2
Male	1.7	1.3	1.2
Measurement	2.0	1.7	1.9
Female	2.1	1.9	1.9
Male	2.4	2.3	2.5
Geometry	1.4	1.3	1.3
Female	1.6	1.6	1.3
Male	1.6	1.5	1.6
Data analysis	2.1	1.6	1.7
Female	2.0	1.8	2.3
Male	2.6	1.9	1.7
Algebra	1.7	1.4	1.1
Female	1.9	1.8	1.2
Male	2.0	1.8	1.5

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G7.11 — Standard errors for grade-12 mathematics means overall and by gender

	Students from sample S1	Students from sample S2	Students from sample S3
Composite	**1.5**	**1.1**	**1.2**
Female	**1.5**	**1.5**	**1.3**
Male	**1.7**	**1.4**	**1.5**
Numeric operations	1.6	1.1	1.1
Female	1.6	1.5	1.3
Male	2.0	1.4	1.4
Measurement	1.8	1.6	1.4
Female	1.8	2.0	2.0
Male	2.1	1.9	1.9
Geometry	1.5	1.2	1.5
Female	1.7	1.7	1.4
Male	1.7	2.0	1.9
Data analysis	1.5	1.3	1.3
Female	1.5	1.6	1.4
Male	1.8	1.7	1.5
Algebra	1.7	1.2	1.5
Female	1.6	1.6	1.6
Male	2.0	1.5	1.9

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

THE NATION'S REPORT CARD naep

Table G7.12 — Standard errors for grade-4 mathematics means by race/ethnicity

	Students from sample S1	Students from sample S2	Students from sample S3
Composite			
Black	**4.6**	**2.4**	**2.4**
American Indian	**2.8**	**4.2**	**3.3**
Asian/Pacific Islander	**3.9**	**6.1**	**5.3**
Hispanic	**2.8**	**2.1**	**1.9**
White	**1.3**	**1.2**	**1.2**
Numeric operations			
Black	5.4	2.6	2.7
American Indian	3.1	5.3	3.6
Asian/Pacific Islander	4.8	6.9	6.5
Hispanic	2.9	2.6	2.2
White	1.5	1.4	1.3
Measurement			
Black	3.3	3.8	3.9
American Indian	3.8	4.9	3.9
Asian/Pacific Islander	5.3	6.0	6.0
Hispanic	3.5	3.1	2.7
White	1.6	1.4	1.7
Geometry			
Black	4.3	3.1	2.0
American Indian	3.1	4.0	4.8
Asian/Pacific Islander	3.9	6.1	5.7
Hispanic	2.6	2.3	1.9
White	1.3	1.4	1.2
Data analysis			
Black	6.1	3.7	3.0
American Indian	3.2	3.7	3.1
Asian/Pacific Islander	5.1	6.7	5.9
Hispanic	3.5	2.3	2.6
White	1.7	1.8	1.5
Algebra			
Black	4.8	2.9	2.6
American Indian	2.7	4.3	3.7
Asian/Pacific Islander	5.9	6.9	6.7
Hispanic	2.9	2.4	2.4
White	1.5	1.6	1.4

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

THE NATION'S REPORT CARD naep

Table G7.13 — Standards errors for grade-8 mathematics means by race/ethnicity

	Students from sample S1	Students from sample S2	Students from sample S3
Composite			
Black	**3.0**	**2.0**	**2.3**
American Indian	**3.3**	**——**	**——**
Asian/Pacific Islander	**5.3**	**4.0**	**5.0**
Hispanic	**3.4**	**2.4**	**1.9**
White	**1.6**	**1.4**	**1.3**
Numeric operations			
Black	3.4	2.4	1.7
American Indian	2.8	——	——
Asian/Pacific Islander	6.5	4.3	5.1
Hispanic	3.0	2.4	2.7
White	1.5	1.4	1.2
Measurement			
Black	4.7	3.2	4.3
American Indian	7.5	——	——
Asian/Pacific Islander	6.4	6.8	7.8
Hispanic	5.5	3.9	2.8
White	2.3	2.0	2.2
Geometry			
Black	3.1	2.1	3.2
American Indian	4.0	——	——
Asian/Pacific Islander	4.1	3.3	5.0
Hispanic	3.7	2.4	1.7
White	1.5	1.6	1.5
Data analysis			
Black	3.7	2.7	2.2
American Indian	4.4	——	——
Asian/Pacific Islander	6.9	5.7	6.2
Hispanic	4.9	3.4	3.0
White	2.4	1.9	2.2
Algebra			
Black	3.2	2.4	2.4
American Indian	4.8	——	——
Asian/Pacific Islander	5.9	6.4	5.5
Hispanic	2.6	2.6	2.1
White	2.0	1.9	1.3

—— Insufficient sample size to permit reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Table G7.14 — Standard errors for grade-12 mathematics means by race/ethnicity

	Students from sample S1	Students from sample S2	Students from sample S3
Composite			
Black	**3.7**	**2.0**	**2.4**
American Indian	**14.1**	**— —**	**— —**
Asian/Pacific Islander	**6.2**	**3.9**	**3.2**
Hispanic	**2.7**	**2.7**	**2.4**
White	**1.3**	**1.1**	**1.4**
Numeric operations			
Black	3.8	1.9	2.1
American Indian	18.7	— —	— —
Asian/Pacific Islander	7.0	4.8	3.5
Hispanic	2.7	2.4	2.5
White	1.5	1.3	1.3
Measurement			
Black	4.5	3.0	3.4
American Indian	21.4	— —	— —
Asian/Pacific Islander	7.5	4.9	3.8
Hispanic	2.8	3.5	3.4
White	1.7	1.8	1.5
Geometry			
Black	4.1	2.6	2.5
American Indian	9.9	— —	— —
Asian/Pacific Islander	6.2	4.8	4.8
Hispanic	3.2	3.7	2.4
White	1.6	1.3	1.6
Data analysis			
Black	3.8	1.8	2.5
American Indian	11.2	— —	— —
Asian/Pacific Islander	6.8	4.6	4.2
Hispanic	3.0	3.0	2.9
White	1.3	1.4	1.4
Algebra			
Black	4.2	2.2	3.9
American Indian	13.0	— —	— —
Asian/Pacific Islander	6.3	4.0	4.1
Hispanic	3.5	3.4	3.2
White	1.6	1.3	1.8

— — Insufficient sample size is to produce reliable estimate.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Acknowledgments

This research report is the product of the work of many talented staff who contributed their considerable knowledge, technical expertise, and creativity to the NAEP 1996 research efforts. The NAEP mathematics and science assessments, from which the data for this study were drawn, were collaborative efforts among the staff from the National Center for Education Statistics (NCES) in the U.S. Department of Education, the National Assessment Governing Board (NAGB), Educational Testing Service (ETS), Westat, Inc., and National Computer Systems (NCS).

This study was funded through NCES, in the Office of Educational Research and Improvement of the U.S. Department of Education. The acting Commissioner of Education Statistics, Gary Phillips, Associate Commissioner, Peggy Carr, and NCES staff members, Janis Brown, Arnold Goldstein, Andrew Kolstad, Laurence Ogle, Holly Spurlock, Suzanne Triplett, and Sheida White worked closely and collegially with the authors to produce and review this report.

The NAEP project at ETS is directed by Stephen Lazer and John Mazzeo. Mathematics test development activities for 1996 were led by Jeff Haberstroh, science test development activities were led by Christine O'Sullivan, and questionnaire development activities were led by Nancy Anderson (formerly of ETS). Sampling and data collection activities were conducted by Westat under the direction of Rene Slobasky, Nancy Caldwell, Keith Rust, and Dianne Walsh. Printing distribution, scoring, and processing activities were conducted by NCS under the direction of Brad Thayer, Patrick Bourgeacq, Mathilde Kennel, Linda Reynolds, and Brent Studer.

Statistical and psychometric activities for NAEP at ETS are directed by Nancy Allen, John Donoghue, and Frank Jenkins. Data analysis activities for NAEP at ETS are directed by John Barone and David Freund. The analyses presented in this report were conceived and conducted by James Carlson, John Donoghue, John J. Ferris, Steven Isham, Frank Jenkins, Bruce Kaplan, Edward Kulick, and John Mazzeo, Xiaohui Wang, and Lois Worthington. The design and production of the report at ETS was overseen by Rod Rudder, while the text and graphics were produced by Barbette Tardugno, Loretta Casalaina, Kelly Gibson, and Sharon Davis-Johnson. Editing work was coordinated by Shari Santapau, with assistance from Al Benderson, Robert Finnegan, Deborah Kline, Barbara Mitchell, and Martha Thompson.

Many thanks are due to the numerous reviewers, both internal and external to NCES and ETS. The comments and critical feedback of the following reviewers are reflected in this report: Nancy Allen, Jay R. Campbell, John Donoghue, Ray Fields, Arnold Goldstein, Stephen Lazer, David Malouf, Charlene Rivera, and Sharif Shakrani.

United States
Department of Education
Washington, DC 20208–5653

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G-17

Standard Mail (B)