Are NAEP Executive Summary Reports Understandable to Policy Makers and Educators?

CSE Technical Report 430

Ronald K. Hambleton and Sharon C. Slater
University of Massachusetts at Amherst

June 1997

Center for the Study of Evaluation
National Center for Research on Evaluation, Standards,
and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported in part under the National Center for Education Statistics Contract No. RS90159001 as administered by the U.S. Department of Education, Office of Educational Research and Improvement.

ARE NAEP EXECUTIVE SUMMARY REPORTS UNDERSTANDABLE TO POLICY MAKERS AND EDUCATORS?[1]

Ronald K. Hambleton and Sharon C. Slater

University of Massachusetts at Amherst

Abstract

This research study is a follow-up to several recent studies conducted on NAEP reports that found policy makers and the media were misinterpreting text, figures, and tables. Our purposes were (a) to investigate the extent to which NAEP Executive Summary Reports are understandable to policy makers and educators, and (b) to the extent that problems are identified, to offer a set of recommendations for improving NAEP reporting practices.

The main finding from this interview study with 59 policy makers and educators is that, in general, these two groups of NAEP report users had considerable difficulty with the presentation of results in the NAEP Executive Summary Report they were given. Misunderstandings and mistakes in reading the NAEP report were common. Many of the persons interviewed (a) had limited prior exposure to NAEP, (b) were unfamiliar with the NAEP reporting scale, and (c) had a limited knowledge of statistics. These shortcomings contributed substantially to the problems encountered in reading the NAEP Executive Summary Report.

Several recommendations are offered for improving the NAEP reports: First, all displays of data should be field tested prior to their use in NAEP Executive Summary Reports. A second recommendation is that NAEP reports for policy makers and educators should be considerably simplified. A third recommendation is that NAEP reports tailored to particular audiences may be needed to improve clarity, understandability, and usefulness.

Background

The main purpose of the National Assessment of Educational Progress (NAEP) is to provide policy makers, educators, and the public with information about what students in the elementary, middle, and high schools know and can do,

[1] The authors are pleased to acknowledge the constructive suggestions of Ray Fields, Mary Frase, Robert Linn, and Howard Wainer on an earlier draft of this report, and of Daniel Koretz on the design of the study.

and to monitor any changes in student achievement over time. In view of the importance of NAEP data for effective educational policy making and for informing the public about the status of education in America as well as the trends in educational achievement over time, considerable statistical and psychometric sophistication (the best that is available in the country) is used in test design, data collection, test data analysis, and scaling (see, for example, Beaton & Johnson, 1992; Johnson, 1992; Mislevy, Johnson, & Muraki, 1992).

Considerably less attention in the NAEP design has been given to the ways in which data are organized and reported to NAEP audiences, which include policy makers, educators, and the public, though important progress has been made (Beaton & Allen, 1992). Item response theory (IRT) scaling, the use of anchor points and performance standards in score interpretations, and plausible values methodology for obtaining score distributions have been important in enhancing NAEP score reporting. Still, concerns about NAEP data reporting have become an issue in recent years and were documented recently by Jaeger (1992), Koretz and Deibert (1993), Linn and Dunbar (1992), and Wainer (1994, 1995a, 1995b). Controversy, also, exists with respect to the proper interpretations of anchor levels and achievement levels (i.e., performance standards), which have become central concepts in NAEP reporting (American College Testing, 1993; Forsyth, 1991; Hambleton & Bourque, 1991; National Academy of Education, 1993; Stufflebeam, Jaeger, & Scriven, 1991).

The designs of tables, figures, and charts to transmit statistical data to enhance their meaningfulness and understandability is a fairly new area of concern in education and psychology (Wainer, 1992; Wainer & Thissen, 1981). There is however an extensive literature that appears relevant to the topic of data reporting in the fields of statistics and graphic design (see, for example, Cleveland, 1985; Henry, 1995). Related to the problem of reporting designs is the topic of reporting scales, which are also intended to facilitate NAEP data reporting (see, Phillips et al., 1993). But to the extent that the scales are confusing to intended audiences, misinterpretations follow, and the value of NAEP for effective policy making is considerably reduced (see Hambleton & Slater, 1994).

There are many potential threats to the validity of NAEP data. The content frameworks may not reflect national curriculum trends. The assessment material used in the NAEP assessments may be flawed in some way, for example, technical inadequacies, failure to match the objectives the materials were

designed to measure, or biases of one kind or another. Problems with the reporting scales are a possibility because of the strong assumptions that must be met in their construction. There is also the potential problem of low student motivation to perform up to ability levels on assessments such as NAEP, which have low consequences for individuals, schools, and districts (Kiplinger & Linn, 1993).

The list of potential threats to the validity of NAEP results is quite long, but considerable effort is expended by the National Center for Education Statistics (NCES), the National Assessment Governing Board (NAGB), by the contractor, Educational Testing Service (ETS), and by others to minimize these threats. Interested readers are referred to special issues of the *Journal of Educational Measurement* and the *Journal of Educational Statistics* in 1992, which feature many articles on the technical aspects of NAEP. Also, voluminous research reports stretching over a 25-year period are available from the NCES.

Statement of the Problem

There remains one threat to the validity of inferences about NAEP results that, to date, has received considerably less attention than others from researchers: the threat due to misunderstandings of the NAEP reports themselves by intended NAEP audiences. These misunderstandings may be due to overly terse and/or confusing text, overly complex or unclear tables and figures, and other characteristics of the reports.

The problem may not only be due to shortcomings in the design of report forms. Often NAEP audiences are not well prepared to handle the wealth of data that a complex NAEP assessment provides. There may even be questions about the facility of NAEP audiences to handle some fairly basic statistical concepts (such as the distinctions between means and medians, and percentages and percentiles) and interpret even fairly simple graphs and tables. In the presence of severe limitations of some members of intended audiences to comprehend even the simplest of statistical concepts, there are limits on how much can be accomplished with even the clearest of reports. Even the harshest critics of national assessment would have to concede that the task of effectively communicating the richness of the NAEP database to various audiences who have limited expertise in handling statistical information and often limited time with which to read NAEP reports is immensely difficult.

But how bad (or good) is the current situation? Do policy makers and educators understand what they are reading about student achievement and changes over time? Do they make reasonable inferences and avoid inappropriate ones? And what do they think about the information they are being given? Is it important to them? Do they have more success reading tables than some of the charts and plots shown in Appendix A? What do they understand and where are their deficiencies and strengths when it comes to NAEP reports? In view of the shortage of available evidence about the extent to which intended NAEP audiences understand and can use the reports provided by NAEP, research on the topic seemed appropriate. A research study could document not only the extent of understanding and use of various reports by different audiences, but the nature of the problems that might exist, so that NAEP reports, as well as many other reports of test results to policy makers and educators, might be improved.

Purposes of the Investigation

Our research study was stimulated by several recent studies conducted on NAEP reports that found that policy makers and the media were misinterpreting some of the texts, figures, and tables (Jaeger, 1992; Koretz & Deibert, 1993; Linn & Dunbar, 1992). Our purposes were (a) to investigate the extent to which NAEP Executive Summary Reports are understandable to policy makers and educators, and (b) to the extent that problems were identified, to offer a set of recommendations for improving reporting practices.

Such a study seemed essential because there is an unevenness in the measurement literature: There are large numbers of studies on a variety of technical topics such as test development, reliability, validity, standard-setting, and proficiency estimation, but relatively few studies on the topic of reporting test score information to communicate effectively with a variety of audiences (for an important exception, see Aschbacher & Herman, 1991). More research is needed to provide a basis for the development of guidelines.

The goal of this initial study on the validity of data interpretations from NAEP reports was modest. No attempt was made to draw representative samples of persons from the population of readers of NAEP reports, and no attempt was made to comprehensively cover the plethora of NAEP reports, tables, figures, and scales. These points will be discussed in more detail in the next section.

Basic Methodology

NAEP Audiences

Originally we had intended to use three audiences in the study: policy makers (e.g., legislators, legislative assistants), educators (e.g. school superintendents, curriculum specialists), and the media. Members of the media (mainly education writers for newspapers) proved to be difficult to recruit for the study. First, there were few newspaper persons available at any particular site (for example, there were only three or four persons who wrote about education for local papers in the Boston area), and therefore considerable travel (and expense) would have been required to locate a sufficient number of persons in the media for the study. Second, several newspaper writers whom we did contact declined our invitations to participate. They said they preferred asking questions to answering them and would not participate in the study. Because of the cost involved in interviewing persons in the media, the difficulty of finding members of the media to participate, and the modest scope of the study, the media were dropped as a major category of participants. We were able to obtain the cooperation of two members of the press and their responses are contained in the Results section.

Also, in order to minimize costs associated with the interviews, participants were chosen from a small number of sites: Boston, and several communities in Massachusetts; Hartford, Connecticut; Washington, DC; Baton Rouge, Louisiana; and Frankfort, Kentucky. Massachusetts and Connecticut were convenient places for the researchers to visit. Several policy makers in Kentucky and Louisiana had shown some interest in the study. Washington, with the availability of large numbers of educational policy makers, was an obvious choice.

The 59 participants in the interviews comprised a broad audience, similar to the intended audience of the NAEP Executive Summary Reports. Appendix B contains the names, locations, and brief job descriptions of participants. We spoke with persons at state departments of education, attorneys, directors of companies, state politicians and legislative assistants, school superintendents, education reporters, and directors of public relations. Many of the people we interviewed were prominent individuals in their fields, and most held advanced degrees.

NAEP Reports, Scales, Tables, and Figures

Hundreds of reports of NAEP results in many subject areas currently exist in the public domain. For the purposes of this study, we chose initially to focus on reports associated with the 1990 NAEP Mathematics Assessment. This seemed to be a reasonable decision for two reasons: First, achievement levels (i.e., performance standards) were used for the first time in data reporting. In view of the controversy surrounding the use of performance standards in NAEP score reporting, the inclusion of reports containing performance standards seemed like a good idea. Second, the 1990 NAEP Mathematics Assessment was the first to report data at the state level. The addition of performance standards and state data results led to the introduction of many new tables, graphs, and explanations, which appeared to increase the cognitive demands on NAEP report readers.

On the basis of a review of several documents including *The State of Mathematics Achievement* (Mullis, Dossey, Owen, & Phillips, 1991) and *The Levels of Mathematics Achievement* (Bourque & Garrison, 1991), plus the corresponding state reports prepared by NCES and NAGB, many exhibits or displays of data were selected and organized into eight homogeneous groups. Our intention had been to administer each group of exhibits or displays to (up to) nine participants. Federal restrictions prohibited the administration of these materials to more than nine persons without federal review and approval of the materials.

We began the study by drawing materials from various 1990 NAEP reports. Two problems were identified fairly quickly in our research. First, some of the 1990 NAEP displays of data we considered using had already been revised and improved for use in the 1992 NAEP reports. It seemed inappropriate to design our study around outdated displays of data. Second, the use of data displays pulled from the contexts in which they appeared in NAEP reports would complicate the data interpretation task and possibly lead to improper inferences about the extent of understanding of NAEP reports on the part of policy makers and educators. The problem was solved by organizing the study around a single, 30-page NAEP report that could be given to policy makers and educators in its complete form.

After all things were considered, the interviews conducted in the study were designed around the Executive Summary of the *NAEP 1992 Mathematics Report Card for the Nation and the States* (Mullis et al., 1993). This particular report was chosen because it was relatively brief and was intended to stand alone for policy

makers and educators. Also, the NAEP Executive Summary Reports are well known and widely distributed (over 100,000 copies of each Executive Summary are produced) to many people working in education or interested in education. Further, we thought that the NAEP Executive Summary Report results, which included both national and state results, would be of interest to the interviewees who were from different areas of the country. Like most executive summaries, this report's format contains tables, charts, and text to present only the major findings of the assessment. For a more in-depth analysis of the NAEP 1992 Mathematics results, readers would need to refer to some of the more comprehensive NAEP reports prepared by NCES. The materials around which the interview was organized are contained in Appendix A.

Our goal in the interviews was to determine just how much of the information reported in the Executive Summary Report was understandable to the intended audiences. We attempted to pinpoint the aspects of reporting that were confusing to readers, and to identify changes in the reporting that the interviewees felt would improve their understanding of the results.

The 1992 NAEP Mathematics Executive Summary Report consists of six sections that highlight the findings from different aspects of the assessment. For each section, interview questions were designed in an attempt to ascertain the kind of information interviewees were obtaining from the report. Interviewees were asked to read a brief section of the report, and then they were questioned on the general meaning of the text or on the specific meaning of certain phrases. Interviewees also examined tables and charts and were asked to interpret some of the numbers and symbols. Throughout the interviews, we encouraged the interviewees to volunteer their opinions or suggestions. This kind of information helped us gain a general sense of what the interviewees felt was helpful or harmful to them when trying to understand statistical information. The interview form is shown in Appendix C. Some initial field-test work was carried out in Amherst, Massachusetts, prior to its use, and then several improvements and extensions to the interview form were made during the course of the interviews.

Results

In this section, the responses to the interview questions will be described. In particular, the incorrect responses and misconceptions that we discovered will be highlighted.

Not all interviewees were asked all questions. In order to keep the typical interview between 45 minutes and an hour (several of the interviews exceeded 90 minutes), each interviewee was questioned on only three sections of the report, with additional sections if time was available. All interviewees responded to two sections, Major Findings (Section 1) and Achievement Levels (Section 3); the majority of interviewees responded to a third section, Overall Mathematics Performance for the States (Section 4).

The number of interviewees questioned on the last two sections (Performance for Demographic Subpopulations [Section 5], and What Students Know and Can Do in Mathematics [Section 6]) is considerably less than the first three sections. Unfortunately, there was rarely sufficient time in the interviews to address these sections of the NAEP report. Also, during the several months of collecting data, a few questions were added to the interview to gain more specific information about how the interviewees were interpreting the material. We sometimes omitted certain questions if an interviewee was particularly knowledgeable or so confused that follow-up questions on the same topic would have been of limited value. For these reason, the number of responses varied quite a bit from question to question, and small differences across categories and questions should not be interpreted because of the small sample size and the selective way in which the questions were assigned to interviewees. The number of responses per question can be seen in Tables 2 to 7. Distribution of interviewees by type of work is provided in Table 1.

Our sample of interviewees was mainly White and included somewhat more females than males (64% to 36%, Table 2 below). The interviewees were from various areas of education (Table 1 below), and we were able to locate two education reporters for the study. All interviewees indicated that they had medium to high interest in national student achievement results. Further, most (90%) were familiar with NAEP in a general way at least, and 64% had read NAEP publications prior to the interview. Therefore, participants in the study were familiar with the kinds of reports used in the interview. In addition, approximately half the sample had taken more than one course in testing and/or statistics (46%); one fourth only had one course; and one fourth had none. It became clear, however, as the study progressed, that many interviewees had forgotten a lot of the statistical and measurement information they had known at one time.

Table 1

Distribution of Interviewees by Job Description

General job description	Number	Percent
State education agency administrators	12	20.3
Department of education consultants	10	16.9
Department of education researchers	7	11.8
Education reporters	2	3.3
Educators/school administrators	8	13.5
Legislators, legislative assistants, and attorneys	7	11.8
National and regional education organization directors and assistants	13	22.0

Table 2

Background Information on the Interviewees

Characteristic	Level	Number	Percent
Race	Black	3	5.1
	Hispanic	1	1.7
	White	55	93.2
Sex	Male	21	35.6
	Female	38	64.4
Interest level in student achievement	High	41	74.5
	Medium	14	25.5
	Low	0	0.0
Number of statistics or testing courses	More than one	27	45.8
	One	16	27.1
	None	16	27.1
Previous knowledge of NAEP	Yes	52	89.7
	No	4	6.9
	Unsure	2	3.4
Read NAEP reports in the past	Yes	38	64.4
	No	17	28.8
	Unsure	4	6.8
Seen NAEP results in newspapers	Yes	25	75.8
	No	5	15.2
	Unsure	3	9.1

Major Findings Section

Nearly all of the interviewees (92%) demonstrated a general understanding of the main points of the text summarizing the major findings of the NAEP Executive Summary Report (see report pages 1 to 4; see Table 3 below), although several interviewees commented that they would have liked more descriptive information (e.g., concrete examples). One of the problems in understanding the text was due to the use of some statistical jargon (e.g., statistical significance, variance). This confused and even intimidated a small number of the interviewees. Several interviewees suggested that a glossary of basic terms would have enhanced the readability of the report. Terms such as Basic, Proficient, Advanced, standard errors, the NAEP scale, etc. could be included in a glossary.

As one example of a problem, the meaning of the phrase "statistically significant" was unclear to many interviewees (42%). We were looking for an understanding that "statistically significant increases" are not just increases due to chance. We discovered that 58% of the interviewees had an idea, or thought that they knew the meaning, but many of the interviewees in this group could not explain what the term meant or why it was used. This was surprising because more than half the interviewees had taken statistics courses. Typical responses to the question "What does statistically significant mean?" were:

> More than a couple of percentage points.
>
> Ten percentage points.
>
> At least five point increase.
>
> More than a handful—you have enough numbers.
>
> Statisticians decide it is significant due to certain criteria.
>
> The results are important.
>
> I wish you hadn't asked me that. I used to know.

The common mistake was to assume "statistically significant differences" were "big and important differences."

Table 3

Distribution of Responses to Questions From the *Major Findings* Section

Question	Response	Frequency	Percent
What is being said about mathematics achievement at the national level?			
	a. Incorrect	5	8.5
	b Performance improved significantly between 1990 and 1992	37	62.7
	c. Improvement occurred at all three grades and in all types of schools	1	1.7
	d. Both b and c	16	27.1
What does statistically significant mean?			
	Correct	34	57.6
	Incorrect	25	42.4
What does "at or above the Basic level" mean?			
	Correct	38	64.4
	Incorrect	21	35.6
What does "considerable variation in performance" mean?			
	Correct	52	88.1
	Incorrect	6	11.9

Several interviewees mentioned that although they realized that certain terms (e.g., standard error, estimation, confidence level) were important to statisticians, these terms were meaningless to them. After years of seeing these terms in reports, they tended to "glaze over" them when they were used in reports, or formed their own "working" definitions such as those offered above for significance levels.

Another phrase that was problematic for some interviewees (36%) was "60% of the students in grades 4, 8, and 12 were at or above the Basic level." Those who misinterpreted thought that 60% of the students were at the Basic level. This misinterpretation was not due to any memory loss, because the interviewees were looking directly at the phrase when we asked about its meaning. We found that about 36% of the interviewees did not realize that this percentage (60%) also included the percentages of students in the higher categories—in this case, Proficient and Advanced. In this example, they thought that "at or above" included only the students who were in the Basic category. This same type of misunderstanding will be seen in two related questions later in the interview.

Achievement Levels Section

This section of the report (report pages 6 to 10; see Table 4 below) included national and state results regarding the achievement levels—Basic, Proficient, and Advanced. Most interviewees (70%) said that the definitions of Basic, Proficient, and Advanced were clear, but that they didn't hold much meaning. The three levels were defined in relation to each other, but were not defined in an absolute sense:

> The Basic level denotes partial mastery of the knowledge and skills fundamental for Proficient work at each grade. Proficient, the central level, represents solid academic performance and demonstrated competence over challenging subject matter. This is the achievement level the Board has determined all students should reach. The Advanced level signifies superior performance beyond Proficient.

Adding concrete examples of the kinds of skill that students at each level could perform or had mastered was suggested to add more meaning to the definitions. Such information is available in the full NAEP reports. Also, several interviewees had problems with the distinct uses of similar terms: Proficient (meaning the level or category) and proficiency (meaning the scaled-scores). If another term had been used for either one, the report would have been less confusing.

Table 1 of the Executive Summary Report (see Appendix A, Table 1) is one of the most important in the report and contains a wealth of information: Results are reported for Grades 4, 8, and 12; for 1990 and 1992; for average proficiency; for each of the performance categories; and for all statistics in the table, standard errors are given. The confusion about the reporting of "at or above" levels mentioned earlier (and this confusion was repeated in Table 7) was seen again in

Table 4

Distributions of Responses to Questions From the *Achievement Levels* Section

Question	Response	Frequency	Percent
Were the definitions of Basic, Proficient, and Advanced clear?			
	Yes	41	69.5
	No	12	20.3
	Unsure	6	10.2
What does the 18% in line 1 mean?			
	Correct	25	47.2
	Incorrect	28	52.8
What does the 1% in line 2 mean?			
	Correct	48	87.3
	Incorrect	7	12.7
What does the 61% in line 1 mean?			
	Correct	17	81.0
	Incorrect	4	19.0
Do you see any indicators of statistical growth?			
	None	2	3.8
	One	17	32.1
	Two	18	40.0
	Three	16	30.2
What are the standard errors?			
	Correct	38	66.7
	Incorrect	19	33.3
How would you use the standard errors?			
	Correct	18	37.5
	Incorrect	30	62.5
What does the “>” sign mean?			
	Correct	35	66.0
	Incorrect	18	34.0
What does the “<” sign mean?			
	Correct	34	68.0
	Incorrect	16	32.0

Table 4 (continued)

Question	Response	Frequency	Percent
Can you use these symbols correctly?			
	Correct	25	49.0
	Incorrect	7	13.7
	Did not attempt	19	37.3
What is your overall impression of the information in Table 1?			
	Clear	3	6.8
	Needs work	35	79.5
	Unreadable	6	13.6
Do you prefer graphs, tables, or text for statistical information?			
	Graphs	24	47.1
	Tables	6	11.8
	Text	5	9.8
	Graphs and tables	8	15.7
	Graphs, tables and text	2	3.9
	No preference	6	11.8
Is the meaning of the numbers in Table 2 clear to you?			
	Clear	40	70.2
	Not clear	17	29.8
What is the meaning of the “248”?			
	Correct	12	80.0
	Incorrect	3	20.0
Explain what is happening in Table 3.			
	a Best schools have shown real improvement	3	8.3
	b. Poorest schools show less improvement, if any	1	2.8
	c. Both a and b	27	75.0
	d. Incorrect	5	14.3

Table 4 (continued)

Question	Response	Frequency	Percent
What is the size of the difference between the best and the poorest schools?			
	Huge	11	73.3
	Sizeable	2	13.3
	Incorrect	2	13.3
In Table 4, what is the Average Proficiency score for your state?			
	Correct	48	100
	Incorrect	0	0
How does your state compare to the other states?			
	Correct	41	95.3
	Incorrect	2	4.7
Were standard errors used to make this comparison?			
	Yes	2	10.0
	No	18	90.0
Was the regional or national information used to make this comparison?			
	Yes	9	42.9
	No	12	57.1
What percent of the Grade 4 students in your state are performing below Basic?			
	Correct	42	100
	Incorrect	0	0
What percent of the Grade 4 students in your state are Proficient?			
	Correct	18	40.9
	Incorrect	26	59.1

Table 1 of the NAEP Executive Summary Report. When asked what the 18% in line 1 of Table 1 meant (18% of Grade 4 students in 1992 were in the Proficient or Advanced categories in mathematics), over half (53%) of the interviewees responded incorrectly. Several of the interviewees simply did not look at the table

closely enough to see the "Percentage of Students At or Above" heading above the levels. Simply removing the line that separates "Percentage of Students At or Above" from "Basic," "Proficient," and "Advanced" may help to avoid this problem. The fact that the categories were arranged from Advanced to Basic complicated the use of the table and the concept of "at or above." In this case, "at or above" meant summing from right to left, which seemed backwards to interviewees when the correct interpretation was given to them.

The problem that interviewees had with Table 1 cannot be corrected that easily. It just did not make sense to interviewees to report the percentages cumulatively. It was confusing that the columns summed to more than 100. Take, for example, the first line in Table 1 of the NAEP executive Summary Report. The percentages of students at or above the Basic level and below the Basic level add to 100; the interviewees expected *all* columns to add to 100. The percentage listed under the heading "Basic" (61%) includes the percentage of students under the heading Proficient" (18%), which in turn includes the percentage of students listed under "Advanced" (2%). This means that 43% of the students were Basic, 16% were Proficient and 2% were Advanced. A majority of interviewees said they would prefer to have the percentages reported for each performance category separately. If they were interested in cumulative percentage, they would rather sum across the column themselves. A common mistake then was to sum the percentages in line 1 of Table 1 and obtain 120%. Then interviewees who made the mistake were stumped.

We explained how to read the table to those who did not understand that the column percentages were cumulative. We then asked a similar question to see if they now understood how to interpret the table. When asked what the 61% in line 1 meant, all of the interviewees were able to correctly respond that 61% of the students were *at or above* the Basic level.

Only a few interviewees (13%) had difficulty with determining what the 1% in line 2 of Table 1 in Appendix A meant. Without a level above Advanced, this percentage represented the exact amount at that level. This is the kind of discrete column reporting that was familiar to and preferred by the interviewees. One Kentucky educator noted that he was confused because the tables looked different from the tables used in his own state. For example, Kentucky's state summary tables report the percentages of students in *each* proficiency category, and standard errors are not used. This was an interesting and important comment

that may have implications for the design of clear and understandable reports. To the extent that policy makers may be familiar with their own state reporting, variations from that, such as the use of cumulative percentages, may be extra confusing. Perhaps the main point is that care in highlighting special features of reports may be necessary to avoid confusion with other reports that educators and policy makers use.

Another problem with the interpretation of this table was confusion about standard errors. One third (33%) of the interviewees did not know the meaning of standard errors; 62% did not understand how to use them. The footnote below the table in the NAEP Executive Summary Report explaining standard error was too filled with statistical jargon to help those who did not understand the concept. (Even several interviewees who understood the meaning of standard errors, or at least said they did, found the footnote a bit complicated.) Also, only a couple of the interviewees who understood standard errors used them to interpret the results. They relied on the symbols indicating significance to determine whether there was a difference from 1990 to 1992. Over 90% of the interviewees suggested moving the standard errors to an appendix for those who might be interested.

One third of the interviewees also had difficulty with the greater-than (>) and less-than (<) symbols used to denote significance. Because of their use and meaning in mathematics, 34% of the interviewees were confused about how to use them in the table. Also, because of their placement in the table (not next to the numbers or percentages that were significantly different, but beside the standard error), over 50% of the interviewees misinterpreted their meaning. For example, several of the interviewees thought that the ">" symbol indicated the direction of error. Using an asterisk instead of greater-than and less-than symbols would be clearer, simply because it is a more familiar symbol for denoting statistical significance. The actual numbers are sufficient to indicate direction.

From the number of mistakes and misinterpretations made in reading this table, it is not surprising that nearly 80% of the interviewees said that this table "needs work." Several interviewees would replace it entirely with something more visual, like a bar graph. Nearly half of the interviewees prefer to see statistical information presented in graphs. Over 90% of the persons we interviewed indicated that they did not have a lot of time to spend interpreting complex tables like these, and a simple graph can be understood relatively quickly. Several

interviewees took an opposite position to the majority: They would prefer receiving a more lengthy report, if it were a bit more clear and easy to understand.

Table 2 of the NAEP Executive Summary Report (see Appendix A, Table 2) was unclear to about 30% of the interviewees. "Cutpoint" and "scale score" are jargon and were the source of the confusion. One interviewee thought that the numbers in Table 2 represented the numbers of students in each category. Regardless of whether or not the interviewees understood the meaning of the numbers in the table, several interviewees wondered why it was included as a separate table in the report. They commented that without examples or descriptions of skills, the numbers meant nothing to them. Suggestions were made to combine these numbers with the definitions of Basic, Proficient, and Advanced or to present the cutpoints with a graphic instead of in a table. No interviewees had any idea of the meaning of the numbers on the NAEP scale, and this information was not contained in the report.

Table 4 of the NAEP Executive Summary Report (see Appendix A, Table 4) gave the interviewees the least trouble of all of the tables in the report. Perhaps one reason was that the interviewers had corrected any misconceptions or misinterpretations the interviewees had with Table 1 which allowed the interviews to proceed with interviewees having an understanding of the "Percentages of Students At or Above." They were all able to locate the Average Proficiency score and the percentage of students in Grade 4 that performed below Basic. Almost all (95%) were able to compare their state's data to the other states, as well. Only 2 of 20 interviewees used the standard errors to make this comparison, and 9 (of 21) used the regional or national information given at the top of this table.

Again, a factor that caused great confusion with the interpretation of numbers in this table was the cumulative column percentages. Interviewees seemed to understand the column headings but they were unable to carry out some simple calculation. Only 18 of 44 or 41% were able to calculate correctly the percentage of 4th graders in their state who were considered Proficient. It was not clear to these interviewees that to determine the exact percentage for the Proficient and Basic columns, the percentage of the next highest category had to be subtracted from the value in the column of interest.

Overall Mathematics Performance for the States Section

In this section of the NAEP Executive Summary Report (report pages 11 to 16; see Table 5 below), state results are ranked and compared. One chart, Figure 1 in the report (see Appendix A, Figure 1), contains every possible pairwise comparison between the states and territories who participated in the study. This chart (referred to by some as the "panty hose chart") was a problem for 41% of the interviewees.

The common mistake made when asked "how many states did significantly better than your state?" was to count the number of states listed to the left of their state at the top of the page. Several interviewees simply laughed (out of nervousness) when they saw this figure and the next one in the report and indicated a desire to move on with the interview. Perhaps the chart was unclear because the shading was poor. Possibly the problem is with the meaning of "statistically significant." As mentioned earlier, 41% of the interviewees did not seem to completely grasp this concept. The other big possibility is that the chart contains a tremendous amount of information, perhaps more than many readers can handle at one time, or handle effectively without clearer directions. One revision might be to simply list each state, and then identify the states performing significantly better, about the same, and significantly worse than that state in columns or rows. More space would be needed in reporting the information, but the information itself would be more clear to users.

Once we explained how to use the figure, nearly all interviewees understood it. It was such an unfamiliar chart format that instruction was necessary for them to understand. The directions given at the top of the chart were not sufficient for the interviewees. In addition to the existing directions, an example of how to read the chart would be very helpful—something like:

> Take Utah, for example. Two states performed significantly better than Utah, 21 states showed no statistically significant difference from Utah, and 20 states performed significantly lower than Utah.

Table 5

Distribution of Responses to Questions From the *Overall Mathematics Performance for the States* Section

Question	Response	Frequency	Percent
How many states did significantly better at Grade 4 than your state?			
	Correct	22	59.4
	Incorrect	15	40.6
How many states did your state significantly outperform at Grade 4?			
	Correct	22	57.9
	Incorrect	16	42.1
How did your state rank in Grade 4 mathematics?			
	Correct	38	100
	Incorrect	0	0
What do the black bands in Figure 2 represent?			
	Correct	22	59.4
	Incorrect	15	40.6
Would the ranking be the same if the 25th percentile points were used instead of the mean?			
	Identical	3	25.0
	No, but similar	9	75.0
Why might ranking states based on percentiles be of interest?			
	Correct	9	81.8
	Incorrect	2	18.2
What is your opinion of the clarity of Figures 1 and 2?			
	Clear	3	17.6
	Somewhat clear	2	11.8
	Confusing	6	35.3
	Very confusing	6	35.3

No one had problems determining how his or her state ranked in Grade 4 mathematics (see Appendix A, Figure 2). All interviewees knew to count down the list to see how the state ranked. The meaning of the black band in the center of the bar for each state, however, was not as easy for everyone to understand. Forty-one percent of those asked did not know. Moving the legend to the top of the chart may help; interviewees simply did not seem to see it in the lower left-hand corner of the page. A small number of interviewees did refer to the footnote at the bottom for an explanation, but this provided little help due to the use of statistical jargon.

We were able to ask only a few interviewees the two questions about ranking on the basis of particular percentiles. Most of those asked (9 of 11) understood that ranking based on percentiles would be slightly different and that this kind of ranking would provide more accurate information about the students at the low end and the high end of mathematics achievement in each state. When questioned about the clarity of these two figures, 12 of the 17 interviewees asked said that the figures were confusing (or very confusing) to them.

What interviewees did like was the map of the United States (not included in this report) showing the states who improved, stayed the same, or declined between the 1990 and 1992 in mathematics achievement. This graphic was easy for persons to understand and use. It seemed to motivate interviewees to dig a bit deeper into the Executive Summary. Unfortunately, we had time to discuss this figure with only eight of the interviewees.

Performance for Demographic Subpopulations Section

Only eight interviewees were questioned about this section of the report (report pages 17 to 21; Table 6 below) because of time constraints. All eight were able to understand Table 5 of the NAEP Executive Summary Report (see Appendix A, Table 5).

The purpose of Figure 6 of the report (see Appendix A, Figure 6) was clear to all eight interviewees, as well. Five of the eight said that the presentation of the information in this figure was also clear to them. However, all mentioned that the shading in the figure was quite poor. Interviewees who were questioned on this section tended to be those who had moved quickly through other sections of the report and, of course, had been instructed on problems they had encountered with

Table 6

Distribution of Responses to Questions on the Performance for *Demographic Subpopulations* Section

Question	Response	Frequency	Percent
At the Grade 12 level, which region of the country has the highest mathematics proficiency?			
	Correct	8	100
	Incorrect	0	0
Which region showed a significant increase in performance from 1990 to 1992?			
	Correct	8	100
	Incorrect	0	0
What is the purpose of Figure 6?			
	Correct	8	100
	Incorrect	0	0
Is the presentation of information in Figure 6 clear?			
	Yes	5	62.5
	No	3	37.5

tables appearing earlier in the report. Without some prior instruction, our belief is that these eight interviewees would not have performed nearly as well as they did.

What Students Know and Can Do in Mathematics Section

Only 11 interviewees were asked the questions in this section of the interview (report pages 22 to 28; Table 7 below). Five of the 11 understood the meaning of anchor levels, but only three (of 11) could explain the difference between anchor levels and achievement levels based on material they read in the report. Six (of 11) found the descriptions of anchor levels helpful. These were the kinds of descriptors that interviewees wanted to have with the definitions of the achievement levels presented earlier in the report.

None of the interviewees had problems with the questions asked about Table 7 in the Executive Summary Report (see Appendix A, Table 7). Again, this finding is most likely due to the instruction of interviewees during the interview, since Tables 1 and 7 in the Executive Summary Report conveyed similar material and were similar in format. In general, Table 8 (see Appendix A, Table 8) was easy for

Table 7

Distribution of Responses to Questions From the *What Students Know and Can Do in Mathematics* Section

Question	Response	Frequency	Percent
What is the meaning of anchor levels?			
	Correct	5	45.5
	Incorrect	6	54.5
What is the difference between anchor and achievement levels?			
	Correct	3	27.2
	Incorrect	8	72.8
Were the descriptions of the anchor levels in Table 7 helpful?			
	Yes	6	54.5
	No	2	18.2
	Unsure	3	27.3
What does Table 7 say about the performance of Grade 12 students in the area of reasoning and problem solving involving geometric relationships, algebra, and functions?			
	Correct	8	88.9
	Incorrect	1	11.1
What percent of Grade 4 students in your state were at a score of 200 or above?			
	Correct	8	88.9
	Incorrect	1	11.1
How does this compare to the Nation and the Northeast?			
	Correct	8	80.0
	Incorrect	2	20.0
What is the significance of the fact that 0% or 1% of the Grade 4 students in your state were at a score of 300 or more?			
	Correct	5	45.5
	Incorrect	6	54.5

the interviewees to understand. The question about Table 8 that more interviewees did miss was not a problem with the way the data were presented, but with the way they were interpreted. When asked what the significance of 0% of the Grade 4 students performing at Level 300, two interviewees said that it was unacceptable, or that the 4th graders were not doing well. They didn't take the time to refer back to Table 7 and see that Level 300 corresponded to skills in geometry and algebra, skills that 4th graders are not expected to know. Had they taken the time to study the tables and the meanings of the anchor levels, they probably would not have made this mistake.

Conclusions and Recommendations

Major Findings

The interviewees in the study seemed very interested and willing to participate. For most of them, reports like the NAEP Executive Summary Reports were regularly received in their offices. They were eager to help us to determine the extent to which these reports were understandable, and to be involved in the improvement of these reports by offering their opinions.

Despite the fact that the interviewees tried hard to understand the report, we found that many of them made fundamental mistakes. Nearly all were able to generally understand the text in the report, though many would have liked to see more descriptive information (e.g., definitions of measurement and statistical jargon, and concrete examples). The problems in understanding the text involved the use of statistical jargon. This confused and even intimidated some of the interviewees. Some mentioned that, although they realized that certain terms were important to statisticians, those terms were meaningless to them. After years of seeing these terms in reports, they tended to "glaze over" them.

The tables were more problematic than the text for most of the interviewees. Although most were able to get a general feeling of what the data in the tables meant, many mistakes were made when we asked the interviewees specific questions. The symbols in the tables (e.g., to denote statistical significance) confused some, and others just chose to disregard them. For example, interviewees often "eyeballed" the numbers to determine whether there was improvement, ignoring the symbols next to the numbers denoting statistical significance.

Improvement to these interviewees often meant a numerical increase of any magnitude from one year to the next.

Consider again Table 1 from the NAEP Executive Summary Report and reproduced in Table 8 below. We will use this table to illustrate many of the problems that arose in the use of tables and graphs. Problems that arose with this table are reflective of problems that arose with any tables using a similar format (such as Table 7, which reported data in relation to anchor levels). Policy makers, educators, and the two members of the media who participated indicated several source of confusion:

1. Interviewees were confused by the reporting of average proficiency scores (few understood the 500-point NAEP scale). Also, proficiency as measured by NAEP and reported on the NAEP scale was confused with the category of "proficient students."

2. They were also baffled by the standard error beside each percentage. These were confusing because (a) they got in the way of reading the percentages, and (b) the footnotes did not clearly explain to the interviewees what a standard error is and how it could be used.

3. The < and > signs were misunderstood or ignored by most interviewees. Even after reading the footnotes, many interviewees indicated that they were still unclear about the meaning.

4. The most confusing point for interviewees was the reporting of students *at or above* each proficiency category. Interviewees interpreted these cumulative percents as the percent of students in *each* proficiency category. Then they were surprised and confused when the sum of percentages across any row in Table 8 did not equal 100%. Contributing to the confusion in Table 8 was the presentation of the categories in the reverse order to that which was expected (i.e., Below Basic, Basic, Proficient, and Advanced). This information as presented required reading from right to left instead of the more common left to right. Perhaps only about 10% of the interviewees were able to make the correct interpretations of the percents in Table 8.

5. Footnotes were not always read and were often misunderstood when they were read.

6. Some interviewees expressed confusion due to variations between the NAEP reports and their own state reports.

Table 8

National Overall Average Mathematics Proficiency and Achievement Levels, Grades 4, 8, and 12

Grade	Assessment Year	Average Proficiency	Percentage of Students at or Above Advanced	Proficient	Basic	Percentage Below Basic
4	1992	218(0.7)>	2(0.3)	18(1.0)>	61(1.0)>	39(1.0)<
	1990	213(0.9)	1(0.4)	13(1.1)	54(1.4)	46(1.4)
8	1992	268(0.9)>	4(0.4)	25(1.0)>	63(1.1)>	37(1.1)<
	1990	263(1.3)	2(0.4)	20(1.1)	58(1.4)	42(1.4)
12	1992	299(0.9)>	2(0.3)	16(0.9)	64(1.2)>	36(1.2)<
	1990	294(1.1)	2(0.3)	13(1.0)	59(1.5)	41(1.5)

> The value for 1992 was significantly higher than the value for 1990 at about the 95 percent confidence level.

< The value for 1992 was significantly lower than the value for 1990 at about the 95 percent confidence level. The standard errors of the estimated percentages and proficiencies appear in parentheses. It can be said with 95 percent confidence that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference.

Table 9 below was prepared to respond to many of the criticisms raised by interviewees in the study about Table 8 (Table 1 in the Executive Summary Report; see Appendix A). Modest field-testing during the study indicated that Table 9 was considerably less confusing. A simplified Table 9 may be more useful to intended audiences for the report, but Table 9 may be inconsistent with the reporting requirements of a statistical agency such as NCES.

Table 9

National Overall Average Mathematics Proficiency and Achievement Levels, Grades 4, 8, and 12

Grade	Assessment Year	Average Proficiency	Percentage of Students Below Basic	Basic	Proficient	Advanced
4	1992	218>	39%	43%	16%	2%
	1990	213	46	41	12	1
8	1992	268>	37%	38%	21%	4%
	1990	263	42	38	18	2
12	1992	299>	36%	48%	14%	2%
	1990	294	41	46	11	2

The symbols ">" and "<" are used to highlight differences in the table that are large enough to be real and *not* due to chance factors such as instability in the information. For example, it can be said that average mathematics performance in Grade 4 in 1992 was higher than in 1990.

Another common problem for the interviewees was reading the charts. In an assessment of national scope, it is often necessary to include quite a bit of information in each chart. This requires the use of some elegant graphical techniques. This also tends to add to the complexity of the charts. Although these charts are impressive in the NAEP Executive Summary Report, to those who could not interpret them, they were intimidating. The unfamiliar chart formats were very difficult for many of the interviewees. Once the charts were explained, interviewees understood them, but many interviewees commented that they either couldn't have figured the charts out on their own, or more commonly, that they simply would not have the time in a typical day to devote to a report requiring so much study.

The footnotes were of little help in explaining the tables and charts. They were often lengthy and contained statistical explanations that the interviewees did not understand. As an example, the following is a footnote that many of the interviewees found particularly confusing:

> . . . The between state comparisons take into account sampling and measurement error and that each state is being compared with every other state. Significance is determined by an application of the Bonferroni procedure based on 946 comparisons by comparing the difference between the two means with four times the square root of the sum of the squared standard error.
>
> (Taken from Figure 1, page 12, of the Executive Summary of the *NAEP 1992 Mathematics Report Card for the Nation and the States*.)

The first sentence of this footnote would have been sufficient for the policy makers and educators we interviewed.

Despite the fact that many of the interviewees made mistakes, their overall reactions to the task were positive. Some were surprised to find that when they took the time to look at the report closely, they could understand more than they expected. Again, most noted that they did not have the time needed in a typical day to scrutinize these reports until they could understand them fully. When we apologized to one legislator about the shortage of time we may have allowed for the task, he noted that he had already spent more time with us than he would have spent on his own with the report.

Of those interviewees who had problems, once we explained some of the tables and statistical concepts to them, they found the results easier to understand. There were a few interviewees who became so frustrated with the report or with themselves that they simply gave up trying to understand it.

Everyone offered helpful and insightful opinions about the report. Some common suggestions were made in these comments about how to make the results in reports like the Executive Summary Report more accessible to those with little statistical background. A comment made by a couple of interviewees was that the report appeared to be "written by statisticians, for statisticians." To remedy this, many suggested removing the statistical jargon. It seems that phrases like "statistically significant" do not hold much meaning for the policy makers and educators we interviewed.

Another suggestion was to simplify the tables by placing the standard errors in an appendix. The lengthy footnotes could also be placed in an appendix for those who are interested. These tended to clutter the appearance of tables. Brief footnotes in layman's terms would be preferred by many interviewees in our study. Also, according to many interviewees, presenting some of the information in simple graphs instead of tables would be better. One reason is that a simple graph can be understood relatively quickly.

It can be seen from some of the comments mentioned above, that most interviewees need to be able to quickly and easily understand reports. They simply do not have much time or are unwilling to spend much time. Some interviewees would even prefer receiving a more lengthy report, if it were just a bit more clear and easy to understand.

Our conclusions and recommendations are limited because of (a) the modest nature of the study (only 59 interviews were conducted), (b) the nonrepresentativeness of the persons interviewed (although it was an interesting and important group of policy makers and educators), (c) noncomparable samples used to assess the clarity and understandability of the six sections of the NAEP report, and (d) the use of only one NAEP report in the study. Still, several conclusions and recommendations seem reasonable to make on the basis of the work that was done: (a) There was a considerable amount of misunderstanding about the results reported in the 1992 NAEP Mathematics Assessment Executive Summary Report among the persons studied; (b) improvements in this

type of report would need to include the preparation of substantially more user-friendly reports with considerably simplified figures and tables; and (c) regardless of the technical skills of the audiences, reports ought to be kept straightforward, short and clear because of the short time persons are likely to have to spend with these executive summaries.

On the basis of the findings from this study, several reporting guidelines for NAEP and state assessments can be offered:

1. Charts, figures, and tables should be understandable without reference to the text. (Readers didn't seem willing to search around the text for interpretations.)

2. Always field-test graphs, figures, and tables on focus groups representing the intended audiences; many important things can be learned from field-testing report forms such as features of reports that may be confusing to readers. (The situation is analogous to field-testing assessment materials prior to their use. No respectable testing agency would ever administer important tests without first field-testing its material. The same guideline should hold for the design of report forms.)

3. Be sure that charts, figures, and tables can be reproduced and reduced without loss of quality. (This is important because interesting results will be copied and distributed and we have all been forced to look at bad copies at one time or another. Correct interpretations, let alone interest, can hardly be expected if the reports are unreadable. Shading is particularly problematic.)

4. Graphs, figures, and tables should be kept relatively simple and straightforward to minimize confusion and shorten the time required by readers to identify the main trends in the data.

5. With respect to NAEP Executive Summary reports, provide an introduction to NAEP and NAEP scales, include a glossary, de-emphasize statistical jargon, simplify tables, charts, and graphs, and use more boxes and graphics to highlight the main findings.

6. With various intended audiences, it may be the case that specially-designed reports are needed for each. For example, with policy makers, reports might need to be short, with the use of bullets to highlight main points such as conclusions. Tables might be straightforward with focus on only the most important conclusions and implications. Technical data (such as standard errors) and technical discussions along with methodological details of the study should be avoided. Keep the focus on conclusions and significance, and keep the report short. Interested readers can be referred to other documents for additional information.

With respect to the last recommendation, one policy maker said to us that when he was young he used to keep NAEP reports on his shelf for some time, certainly for many years. The results impressed people and, because of their bulk, they filled up his shelves. But after several years, he felt it acceptable to throw them away. Now, he said, he is older and so he skims the reports and throws them away immediately! The challenge for NCES and other agencies reporting assessment results is to give policy makers a reason to keep the reports and to use them.

NAEP reports, in principle, provide policy makers, educators, education writers, and the public with valuable information. But the burden is on the reporting agency to ensure that the reporting scales used are meaningful to the intended audiences and that the reported scores are valid for this recommended use. At the same time, reporting agencies need to focus considerable attention on the way in which scores are reported to minimize confusion as well as misinterpretation, and to maximize the likelihood that the intended interpretations are made. This will require the adoption and implementation of a set of guidelines for reporting that include the field-testing of all reports to ensure that the reports are being interpreted fully and correctly. Special attention will need to be given to the use of figures and tables, which can convey substantial amounts of data clearly if they are properly designed. "Properly designed" means that they are clear to the audiences for whom they are intended.

The recently published *Adult Literacy Study* (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993), conducted by NCES, Westat, and the Educational Testing Service, appears to have benefitted from some of the earlier evaluations of NAEP reporting and provides some excellent examples of data reporting. A broad program of research involving measurement specialists, graphic design specialists (see, for example, Cleveland, 1985), and focus groups representing intended audiences for reports is very much in order to build on some of the successes in reporting represented in the *Adult Literacy Study* and some of the useful findings reported by Jaeger (1992), Koretz and Deibert (1993), Wainer (1994, 1995a, 1995b), and others. Ways need to be found to balance statistical rigor and accuracy in reporting with the informational needs, time constraints, and quantitative literacy of intended audiences.

References

American College Testing. (1993). *Setting achievement levels on the 1992 National Assessment of Educational Progress in mathematics, reading, and writing: A technical report on reliability and validity*. Iowa City, IA: Author.

Aschbacher, P. R., & Herman, J. L. (1991). *Guidelines for effective score reporting* (CSE Tech. Rep. No. 326). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Beaton, A. E., & Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics*, *17*, 191-204.

Beaton, A. E., & Johnson, E. G. (1992). Overview of the scaling methodology used in the National Assessment. *Journal of Educational Measurement, 29*, 163-176.

Bourque, M. L., & Garrison, H. H. (1991). *The levels of mathematics achievement. Volume I: National and state summaries*. Washington, DC: National Assessment Governing Board.

Cleveland, W. S. (1985). *The elements of graphing data*. Monterey, CA: Wadsworth.

Forsyth, R. A. (1991). Do NAEP scales yield valid criterion-referenced interpretations? *Educational Measurement: Issues and Practice, 10*(3), 3-9, 16.

Hambleton, R. K., & Bourque, M. L. (1991). *The levels of mathematics achievement: Initial performance standards for the 1990 NAEP Mathematics Assessment*. Washington, DC: National Assessment Governing Board.

Hambleton, R. K., & Slater, S. (1994, April). *Using performance standards to report national and state assessment data: Are the reports understandable and how can they be improved?* Paper presented at the Joint Conference on Standard-Setting for Large-Scale Assessments, Washington, DC.

Henry, G. T. (1995). *Graphing data: Techniques for display and analysis*. Thousand Oaks, CA: Sage Publications.

Jaeger, R. (1992). General issues in reporting of the NAEP trial state assessment results. In R. Glaser & R. Linn (Eds.), *Assessing student achievement in the states* (pp. 107-109). Stanford, CA: National Academy of Education.

Johnson, E. G. (1992). The design of the National Assessment of Educational Progress. *Journal of Educational Measurement, 29*, 95-110.

Kiplinger, V. L., & Linn, R. L. (1993). *Raising the stakes of test administration: The impact on student performance on NAEP* (CSE Tech. Rep. No. 360). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey*. Washington, DC: U.S. Government Printing Office.

Koretz, D., & Deibert, E. (1993). *Interpretations of National Assessment of Educational Progress (NAEP) anchor points and achievement levels by the print media in 1991*. Santa Monica, CA: RAND.

Linn, R. L., & Dunbar, S. B. (1992). Issues in the design and reporting of the National Assessment of Educational Progress. *Journal of Educational Measurement, 29*, 177-194.

Mislevy, R. J., Johnson, E. G., & Muraki, E. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics, 17*, 131-154.

Mullis, I. V. S., Dossey, J. A., Owen, E. H., & Phillips, G. W. (1991). *The state of mathematics achievement: NAEP's 1990 assessment of the nation and the trial assessment of the states*. Washington, DC: National Center for Education Statistics.

Mullis, I. V. S., Dossey, J. A., Owen, E. H., & Phillips, G. W. (1993). *Executive summary of the NAEP 1992 mathematics report card for the nation and the states*. Washington, DC: Department of Education.

National Academy of Education. (1993). *A report of the National Academy of Education Panel on the evaluation of the NAEP trial state assessment: An evaluation of the 1992 achievement levels*. Stanford, CA: Stanford University, National Academy of Education.

Phillips, G. W., Mullis, I. V. S., Bourque, M. L., Williams, P. L., Hambleton, R. K., Owen, E. H., & Barton, P. E. (1993). *Interpreting NAEP scales*. Washington, DC: Department of Education.

Stufflebeam, D. L., Jaeger, R. M., & Scriven, M. (1991). *Summative evaluation of the National Assessment Governing Board's inaugural effort to set achievement levels on the National Assessment of Educational Progress*. Kalamazoo, MI: Western Michigan University.

Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher, 21*(1), 14-23.

Wainer, H. (1994). *Using trilinear plots for NAEP state data* (Research Rep. No. 94-6). Princeton, NJ: Educational Testing Service.

Wainer, H. (1995a). *Depicting error* (Research Rep. No. 95-2). Princeton, NJ: Educational Testing Service.

Wainer, H. (1995b). *A study of display method for NAEP results: I. Tables* (Research Rep. No. 95-1). Princeton, NJ: Educational Testing Service.

Wainer, H., & Thissen, D. (1981). Graphical data analysis. *Annual Review of Psychology, 32*, 191-241.

Appendix A

Key Tables and Figures From the Executive Summary of the NAEP 1992 Mathematics Report Card for the Nation and the States

TABLE 1 National Overall Average Mathematics Proficiency and Achievement Levels, Grades 4, 8, and 12

			Percentage of Students At or Above			
Grades	**Assessment Years**	**Average Proficiency**	**Advanced**	**Proficient**	**Basic**	**Percentage Below Basic**
4	**1992**	218(0.7)>	2(0.3)	18(1.0)>	61(1.0)>	39(1.0)<
	1990	213(0.9)	1(0.4)	13(1.1)	54(1.4)	46(1.4)
8	**1992**	268(0.9)>	4(0.4)	25(1.0)>	63(1.1)>	37(1.1)<
	1990	263(1.3)	2(0.4)	20(1.1)	58(1.4)	42(1.4)
12	**1992**	299(0.9)>	2(0.3)	16(0.9)	64(1.2)>	36(1.2)<
	1990	294(1.1)	2(0.3)	13(1.0)	59(1.5)	41(1.5)

> The value for 1992 was significantly higher than the value for 1990 at about the 95 percent confidence level. < The value for 1992 was significantly lower than the value for 1990 at about the 95 percent confidence level. The standard errors of the estimated percentages and proficiencies appear in parentheses. It can be said with 95 percent confidence that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

TABLE 2 Mathematics Proficiency (Scale-Score Cutpoint) Corresponding to Each Achievement Level, Grades 4, 8, and 12

Grades	Advanced	Proficient	Basic
4	280	248	211
8	331	294	256
12	366	334	287

TABLE 3 Average Mathematics Proficiency and Achievement Levels for the Top One-Third of the Schools and the Bottom One-Third of the Schools, Grades 4, 8, and 12

				Percentage of Students At or Above			
Grades	**Assessment Years**	**Percent of Students**	**Average Proficiency**	**Advanced**	**Proficient**	**Basic**	**Percentage Below Basic**
Grades 4							
Top One-Third Schools	**1992**	34(2.8)	237(0.8)>	5(0.8)	34(1.5)>	84(1.0)>	16(1.0)<
	1990	34(3.9)	229(1.4)	3(1.1)	25(2.6)	76(1.8)	24(1.8)
Bottom One-Third Schools	**1992**	29(2.1)	196(1.2)	0(0.1)	4(0.5)	32(1.5)	68(1.5)
	1990	30(3.4)	194(1.7)	0(0.2)	4(0.9)	29(2.5)	71(2.5)
Grades 8							
Top One-Third Schools	**1992**	29(3.1)	289(1.3)>	8(1.1)	45(2.0)>	86(1.5)>	14(1.5)<
	1990	30(4.4)	280(1.2)	5(1.0)	35(2.0)	78(1.7)	22(1.7)
Bottom One-Third Schools	**1992**	32(1.8)	245(0.9)	0(0.3)	8(0.8)	37(1.4)	63(1.4)
	1990	34(3.9)	244(1.8)	0(0.3)	8(1.3)	36(2.0)	64(2.0)
Grades 12							
Top One-Third Schools	**1992**	35(3.1)	316(1.1)>	4(0.7)	29(1.5)	82(1.3)>	18(1.3)<
	1990	34(5.0)	310(1.2)	4(0.9)	23(2.3)	77(1.8)	23(1.8)
Bottom One-Third Schools	**1992**	27(2.2)	279(1.0)>	0(0.2)	5(0.9)	40(1.6)	60(1.6)
	1990	26(3.3)	274(1.5)	0(0.2)	3(0.9)	35(2.7)	65(2.7)

> The value for 1992 was significantly higher than the value for 1990 at about the 95 percent confidence level. < The value for 1992 was significantly lower than the value for 1990 at about the 95 percent confidence level. The standard errors of the estimated percentages and proficiencies appear in parentheses. It can be said with 95 percent confidence for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. When the proportion of students is either 0 percent or 100 percent, the standard error is inestimable. However, percentages 99.5 percent and greater were rounded to 100 percent and percentages 0.5 percent or less were rounded to 0 percent.

TABLE 4 **Overall Average Mathematics Proficiency and Achievement Levels**

PUBLIC SCHOOLS	Grade 4 - 1992				
	Average Proficiency	Percentage of Students At or Above Advanced	Percentage of Students At or Above Proficient	Percentage of Students At or Above Basic	Percentage of Students Below Basic
NATION	217 (0.8)	2 (0.3)	18 (1.1)	59 (1.1)	41 (1.1)
Northeast	223 (2.1)	3 (0.8)	23 (2.9)	64 (3.0)	36 (3.0)
Southeast	209 (1.9)	1 (0.4)	11 (1.4)	48 (2.5)	52 (2.5)
Central	222 (2.2)	2 (0.6)	20 (2.1)	66 (3.2)	34 (3.2)
West	217 (1.6)	2 (0.7)	17 (2.1)	59 (2.2)	41 (2.2)
STATES					
Alabama	207 (1.6)	1 (0.2)	10 (1.3)	45 (2.2)	55 (2.2)
Arizona	214 (1.1)	1 (0.3)	13 (0.9)	55 (1.7)	45 (1.7)
Arkansas	209 (0.9)	1 (0.2)	10 (0.8)	49 (1.3)	51 (1.3)
California	207 (1.6)	2 (0.5)	13 (1.2)	48 (2.0)	52 (2.0)
Colorado	220 (1.0)	2 (0.4)	18 (1.1)	62 (1.4)	38 (1.4)
Connecticut	226 (1.2)	4 (0.6)	25 (1.4)	69 (1.5)	31 (1.5)
Delaware	217 (0.8)	2 (0.4)	17 (0.8)	56 (1.0)	44 (1.0)
Dist. Columbia	191 (0.5)	1 (0.2)	6 (0.3)	25 (1.0)	75 (1.0)
Florida	212 (1.5)	2 (0.4)	14 (1.4)	53 (2.0)	47 (2.0)
Georgia	214 (1.3)	2 (0.4)	16 (1.2)	55 (1.7)	45 (1.7)
Hawaii	213 (1.3)	2 (0.4)	15 (1.0)	54 (1.8)	46 (1.8)
Idaho	220 (1.0)	1 (0.3)	16 (1.1)	64 (1.7)	36 (1.7)
Indiana	220 (1.1)	2 (0.3)	16 (1.1)	62 (1.6)	38 (1.6)
Iowa	229 (1.1)	3 (0.5)	27 (1.3)	74 (1.4)	26 (1.4)
Kentucky	214 (1.0)	1 (0.5)	13 (1.1)	53 (1.5)	47 (1.5)
Louisiana	203 (1.4)	1 (0.2)	8 (0.8)	41 (2.0)	59 (2.0)
Maine	231 (1.0)	3 (0.6)	28 (1.5)	76 (1.3)	24 (1.3)
Maryland	216 (1.3)	3 (0.4)	19 (1.2)	57 (1.6)	43 (1.6)
Massachusetts	226 (1.2)	3 (0.5)	24 (1.5)	70 (1.6)	30 (1.6)
Michigan	219 (1.8)	2 (0.5)	19 (1.7)	62 (2.2)	38 (2.2)
Minnesota	227 (0.9)	3 (0.5)	27 (1.2)	72 (1.4)	28 (1.4)
Mississippi	200 (1.1)	0 (0.1)	7 (0.7)	37 (1.3)	63 (1.3)
Missouri	221 (1.2)	2 (0.3)	19 (1.3)	64 (1.6)	36 (1.6)
Nebraska	224 (1.3)	3 (0.5)	23 (1.7)	68 (1.8)	32 (1.8)
New Hampshire	229 (1.2)	3 (0.6)	26 (1.7)	74 (1.6)	26 (1.6)
New Jersey	226 (1.5)	3 (0.7)	25 (1.6)	70 (2.1)	30 (2.1)
New Mexico	212 (1.5)	1 (0.4)	11 (1.3)	52 (1.9)	48 (1.9)
New York	217 (1.3)	2 (0.3)	17 (1.3)	59 (1.9)	41 (1.9)
North Carolina	211 (1.1)	2 (0.4)	13 (0.9)	52 (1.6)	48 (1.6)
North Dakota	228 (0.8)	2 (0.3)	23 (1.1)	74 (1.2)	26 (1.2)
Ohio	217 (1.2)	2 (0.3)	17 (1.1)	59 (1.7)	41 (1.7)
Oklahoma	219 (1.0)	1 (0.4)	14 (1.1)	62 (1.6)	38 (1.6)
Pennsylvania	223 (1.4)	3 (0.5)	23 (1.5)	66 (1.9)	34 (1.9)
Rhode Island	214 (1.6)	2 (0.4)	14 (1.2)	56 (2.2)	44 (2.2)
South Carolina	211 (1.1)	1 (0.3)	13 (1.1)	49 (1.5)	51 (1.5)
Tennessee	209 (1.4)	1 (0.2)	10 (1.0)	49 (2.1)	51 (2.1)
Texas	217 (1.3)	2 (0.5)	16 (1.3)	58 (1.7)	42 (1.7)
Utah	223 (1.0)	2 (0.3)	20 (1.1)	67 (1.6)	33 (1.6)
Virginia	220 (1.3)	3 (0.7)	19 (1.6)	60 (1.4)	40 (1.4)
West Virginia	214 (1.1)	1 (0.3)	13 (1.0)	54 (1.6)	46 (1.6)
Wisconsin	228 (1.1)	3 (0.5)	25 (1.4)	72 (1.3)	28 (1.3)
Wyoming	224 (1.0)	2 (0.3)	19 (1.2)	70 (1.4)	30 (1.4)
TERRITORY					
Guam	191 (0.8)	0 (0.1)	5 (0.5)	28 (1.2)	72 (1.2)

The standard errors of the estimated percentages and proficiencies appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. When the proportion of students is either 0 percent or 100 percent, the standard error is inestimable. However, percentages 99.5 percent and greater were rounded to 100 percent and percentages less than 0.5 percent were rounded to 0 percent.

TABLE 5 Average Mathematics Proficiency by Gender, Race/Ethnicity, Type of Community, and Region

	Assessment Years	Grade 4	Grade 8	Grade 12
Male	1992	220(0.8)>	267(1.1)>	301(1.1)>
	1990	214(1.2)	263(1.6)	297(1.4)
Female	1992	217(1.0)>	268(1.0)>	297(1.0)>
	1990	212(1.1)	262(1.3)	292(1.3)
White	1992	227(0.9)>	277(1.0)>	305(0.9)>
	1990	220(1.1)	270(1.4)	300(1.2)
Black	1992	192(1.3)	237(1.4)	275(1.7)>
	1990	189(1.8)	238(2.7)	268(1.9)
Hispanic	1992	201(1.4)	246(1.2)	283(1.8)>
	1990	198(2.0)	244(2.8)	276(2.8)
Asian/Pacific Islander	1992	231(2.4)	288(5.5)	315(3.5)
	1990	228(3.5)	279(4.8)!	311(5.2)
American Indian	1992	209(3.2)	254(2.8)	281(9.0)
	1990	208(3.9)	246(9.4)	288(10.2)!
Advantaged Urban	1992	237(2.1)	288(3.6)	316(2.6)
	1990	231(3.0)	280(3.2)	306(6.2)
Disadvantaged Urban	1992	193(2.8)	238(2.6)<	279(2.4)
	1990	195(3.0)	249(3.8)!	276(6.0)
Extreme Rural	1992	216(3.6)	267(4.6)	293(1.9)
	1990	214(4.9)	257(4.4)	293(3.3)
Other	1992	219(0.9)>	268(1.1)>	300(0.9)>
	1990	213(1.1)	262(1.7)	295(1.3)
Northeast	1992	223(2.0)>	269(2.7)	302(1.5)
	1990	215(2.9)	270(2.8)	300(2.3)
Southeast	1992	210(1.6)>	260(1.4)	291(1.4)>
	1990	205(2.1)	255(2.5)	284(2.2)
Central	1992	223(1.9)>	274(1.9)>	303(1.8)
	1990	216(1.7)	266(2.3)	297(2.6)
West	1992	218(1.5)	268(2.0)>	298(1.7)
	1990	216(2.4)	261(2.6)	294(2.6)

> The value for 1992 was significantly higher than the value for 1990 at about the 95 percent confidence level. < The value for 1992 was significantly lower than the value for 1990 at about the 95 percent confidence level. ! Interpret with caution -- the nature of the sample does not allow accurate determination of the variability of this estimated statistic. The standard errors of the estimated proficiencies appear in parentheses. It can be said with 95 percent confidence for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

TABLE 7 National Overall Average Mathematics Proficiency and Anchor Levels, Grades 4, 8, and 12

		Assessment Years	Grade 4	Grade 8	Grade 12
Average Proficiency		**1992**	218(0.7)>	268(0.9)>	299(0.9)>
		1990	213(0.9)	263(1.3)	294(1.1)
Level	**Description**		**Percentage of Students at or Above**		
200	**Addition and Subtraction, and Simple Problem Solving with Whole Numbers**	**1992**	72(0.9)>	97(0.4)	100(0.1)
		1990	67(1.4)	95(0.7)	100(0.2)
250	**Multiplication and Division, Simple Measurement, and Two-Step Problem Solving**	**1992**	17(0.8)>	68(1.0)	91(0.5)>
		1990	12(1.1)	65(1.4)	88(0.9)
300	**Reasoning and Problem Solving Involving Fractions, Decimals, Percents, and Elementary Concepts in Geometry, Statistics, and Algebra**	**1992**	0(0.1)	20(0.9)>	50(1.2)>
		1990	0(0.1)	15(1.0)	45(1.4)
350	**Reasoning and Problem Solving Involving Geometric Relationships, Algebra, and Functions**	**1992**	0(0.0)	1(0.2)	6(0.5)
		1990	0(0.0)	0(0.2)	5(0.8)

> The value for 1992 was significantly higher than the value for 1990 at about the 95 percent confidence level. < The value for 1992 was significantly lower than the value for 1990 at about the 95 percent confidence level. The standard errors of the estimated percentages and proficiencies appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. When the proportion of students is either 0 percent or 100 percent, the standard error is inestimable. However, percentages 99.5 percent and greater were rounded to 100 percent and percentages 0.5 percent or less were rounded to 0 percent.

TABLE 8 **Overall Average Mathematics Proficiency and Anchor Levels**

PUBLIC SCHOOLS	Grade 4 - 1992				
	Average Proficiency	Percentage of Students At or Above Level 200	Percentage of Students At or Above Level 250	Percentage of Students At or Above Level 300	Percentage of Students At or Above Level 350
NATION	217 (0.8)	72 (1.0)	16 (0.9)	0 (0.1)	0 (0.0)
Northeast	223 (2.1)	75 (2.5)	22 (2.7)	1 (0.3)	0 (0.0)
Southeast	209 (1.9)	61 (2.4)	10 (1.6)	0 (0.2)	0 (0.0)
Central	222 (2.2)	77 (2.9)	19 (2.0)	0 (0.1)	0 (0.0)
West	217 (1.6)	70 (1.9)	15 (2.0)	0 (0.3)	0 (0.0)
STATES					
Alabama	207 (1.6)	58 (2.1)	9 (1.1)	0 (0.0)	0 (0.0)
Arizona	214 (1.1)	68 (1.5)	12 (0.9)	0 (0.1)	0 (0.0)
Arkansas	209 (0.9)	62 (1.4)	9 (0.7)	0 (0.0)	0 (0.0)
California	207 (1.6)	60 (2.0)	11 (1.1)	0 (0.1)	0 (0.0)
Colorado	220 (1.0)	75 (1.2)	17 (1.0)	0 (0.1)	0 (0.0)
Connecticut	226 (1.2)	79 (1.3)	23 (1.4)	1 (0.3)	0 (0.0)
Delaware	217 (0.8)	69 (1.2)	15 (1.0)	0 (0.1)	0 (0.0)
Dist. Columbia	191 (0.5)	37 (1.5)	5 (0.3)	0 (0.1)	0 (0.0)
Florida	212 (1.5)	66 (1.9)	12 (1.2)	0 (0.2)	0 (0.0)
Georgia	214 (1.3)	67 (1.6)	14 (1.1)	0 (0.1)	0 (0.0)
Hawaii	213 (1.3)	65 (1.6)	14 (0.9)	0 (0.1)	0 (0.0)
Idaho	220 (1.0)	77 (1.6)	14 (1.0)	0 (0.1)	0 (0.0)
Indiana	220 (1.1)	75 (1.4)	14 (1.0)	0 (0.1)	0 (0.0)
Iowa	229 (1.1)	84 (1.1)	24 (1.1)	0 (0.1)	0 (0.0)
Kentucky	214 (1.0)	67 (1.4)	12 (1.0)	0 (0.1)	0 (0.0)
Louisiana	203 (1.4)	54 (1.9)	7 (0.8)	0 (0.1)	0 (0.0)
Maine	231 (1.0)	86 (1.0)	26 (1.5)	1 (0.2)	0 (0.0)
Maryland	216 (1.3)	67 (1.5)	17 (1.2)	0 (0.2)	0 (0.0)
Massachusetts	226 (1.2)	80 (1.1)	22 (1.4)	0 (0.2)	0 (0.0)
Michigan	219 (1.8)	73 (2.0)	17 (1.6)	0 (0.2)	0 (0.0)
Minnesota	227 (0.9)	81 (1.2)	24 (1.1)	0 (0.1)	0 (0.0)
Mississippi	200 (1.1)	50 (1.6)	6 (0.6)	0 (0.1)	0 (0.0)
Missouri	221 (1.2)	76 (1.5)	17 (1.2)	0 (0.1)	0 (0.0)
Nebraska	224 (1.3)	78 (1.5)	20 (1.6)	0 (0.2)	0 (0.0)
New Hampshire	229 (1.2)	84 (1.2)	23 (1.6)	0 (0.2)	0 (0.0)
New Jersey	226 (1.5)	80 (1.8)	23 (1.6)	0 (0.2)	0 (0.0)
New Mexico	212 (1.5)	65 (2.1)	10 (1.3)	0 (0.1)	0 (0.0)
New York	217 (1.3)	71 (1.5)	16 (1.3)	0 (0.2)	0 (0.0)
North Carolina	211 (1.1)	64 (1.6)	12 (0.8)	0 (0.1)	0 (0.0)
North Dakota	228 (0.8)	85 (0.9)	21 (1.1)	0 (0.1)	0 (0.0)
Ohio	217 (1.2)	71 (1.5)	15 (1.1)	0 (0.1)	0 (0.0)
Oklahoma	219 (1.0)	76 (1.5)	13 (1.0)	0 (0.1)	0 (0.0)
Pennsylvania	223 (1.4)	77 (1.5)	20 (1.4)	0 (0.2)	0 (0.0)
Rhode Island	214 (1.6)	68 (1.8)	12 (1.1)	0 (0.1)	0 (0.0)
South Carolina	211 (1.1)	63 (1.3)	12 (1.1)	0 (0.1)	0 (0.0)
Tennessee	209 (1.4)	63 (1.9)	9 (1.0)	0 (0.1)	0 (0.0)
Texas	217 (1.3)	71 (1.8)	14 (1.2)	0 (0.1)	0 (0.0)
Utah	223 (1.0)	79 (1.2)	18 (1.0)	0 (0.1)	0 (0.0)
Virginia	220 (1.3)	73 (1.5)	18 (1.6)	1 (0.3)	0 (0.0)
West Virginia	214 (1.1)	68 (1.6)	11 (0.9)	0 (0.1)	0 (0.0)
Wisconsin	228 (1.1)	83 (1.2)	23 (1.4)	0 (0.2)	0 (0.0)
Wyoming	224 (1.0)	82 (1.2)	17 (1.2)	0 (0.1)	0 (0.0)
TERRITORY					
Guam	191 (0.8)	40 (1.2)	4 (0.5)	0 (0.0)	0 (0.0)

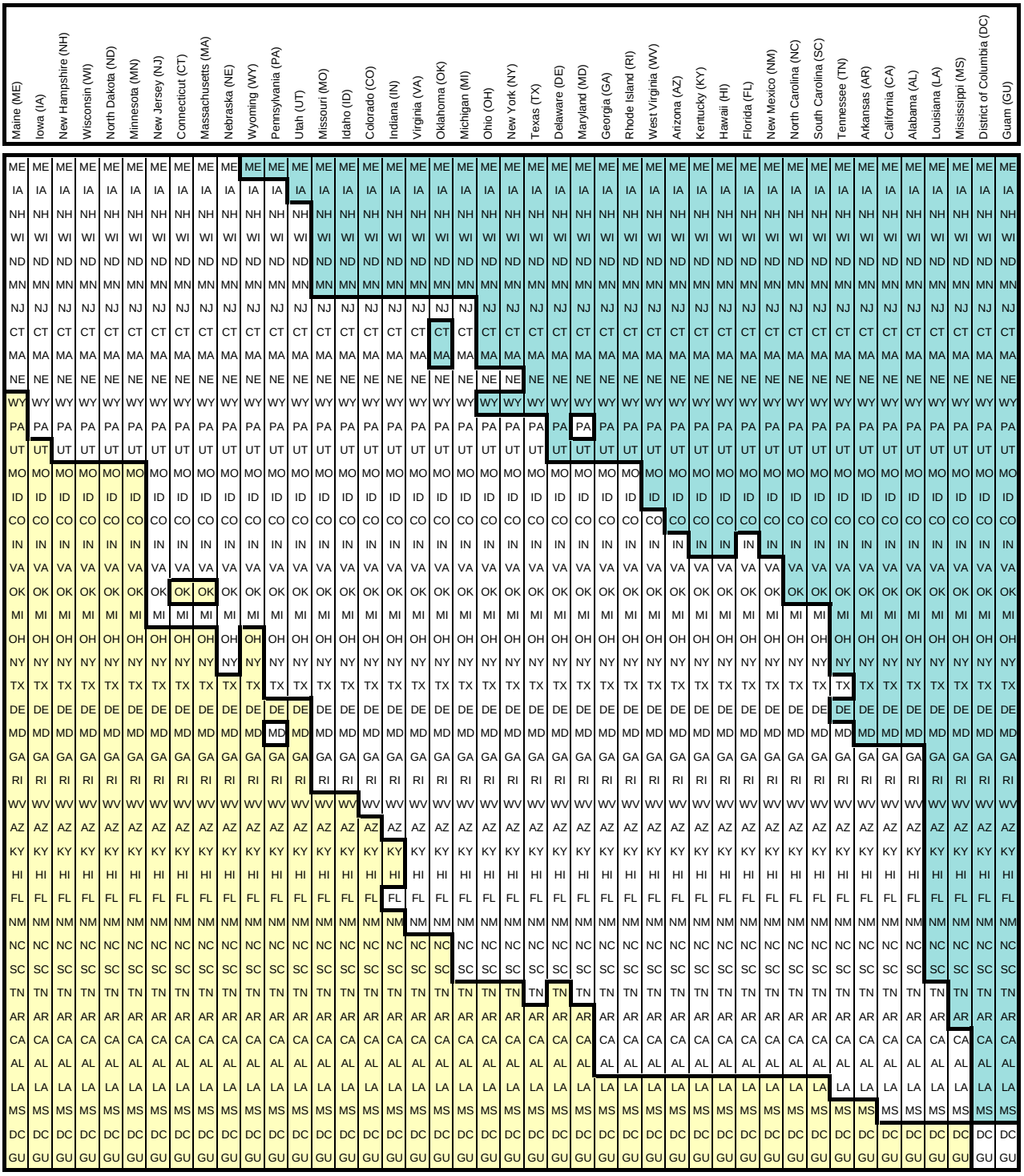
The standard errors of the estimated percentages and proficiencies appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. When the proportion of students is either 0 percent or 100 percent, the standard error is inestimable. However, percentages 99.5 percent and greater were rounded to 100 percent and percentages less than 0.5 percent were rounded to 0 percent.

FIGURE 1

Comparisons of Overall Mathematics Average Proficiency
1992 Grade 4

THE NATION'S REPORT CARD

INSTRUCTIONS: Read *down* the column directly under a state name listed in the heading at the top of the chart. Match the shading intensity surrounding a state postal abbreviation to the key below to determine whether the average mathematics performance of this state is higher than, the same as, or lower than the state in the column heading.



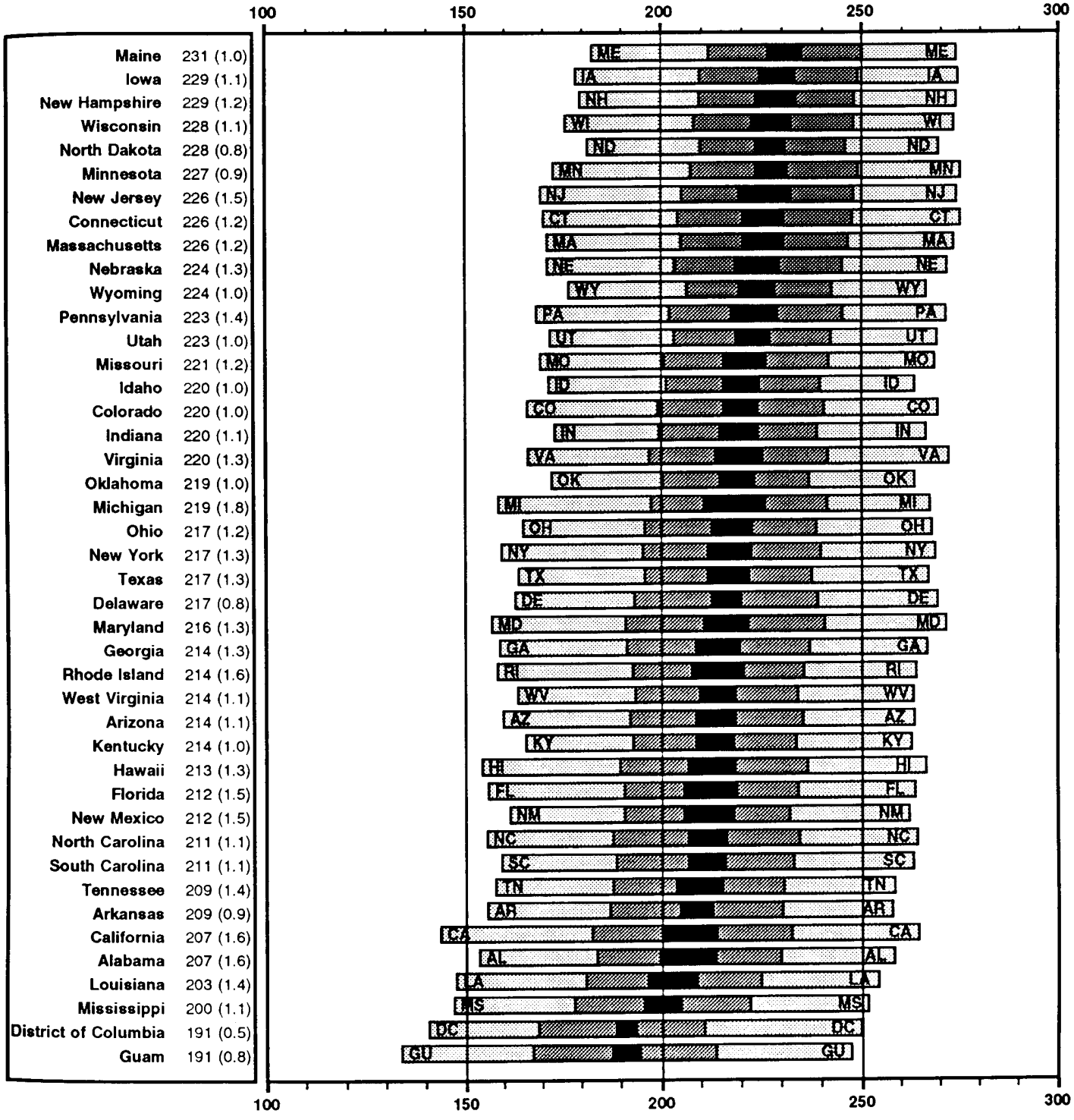
- State has statistically significantly higher average proficiency than the state listed at the top of the chart.
- No statistically significant difference from the state listed at the top of the chart.
- State has statistically significantly lower average proficiency than the state listed at the top of the chart.

The between state comparisons take into account sampling and measurement error and that each state is being compared with every other state. Significance is determined by an application of the Bonferroni procedure based on 946 comprisons by comparing the difference between the two means with four times the square root of the sum of the squared standard errors.

FIGURE 2

Distribution of Overall Mathematics Proficiency Organized by Average Proficiency
1992 Grade 4

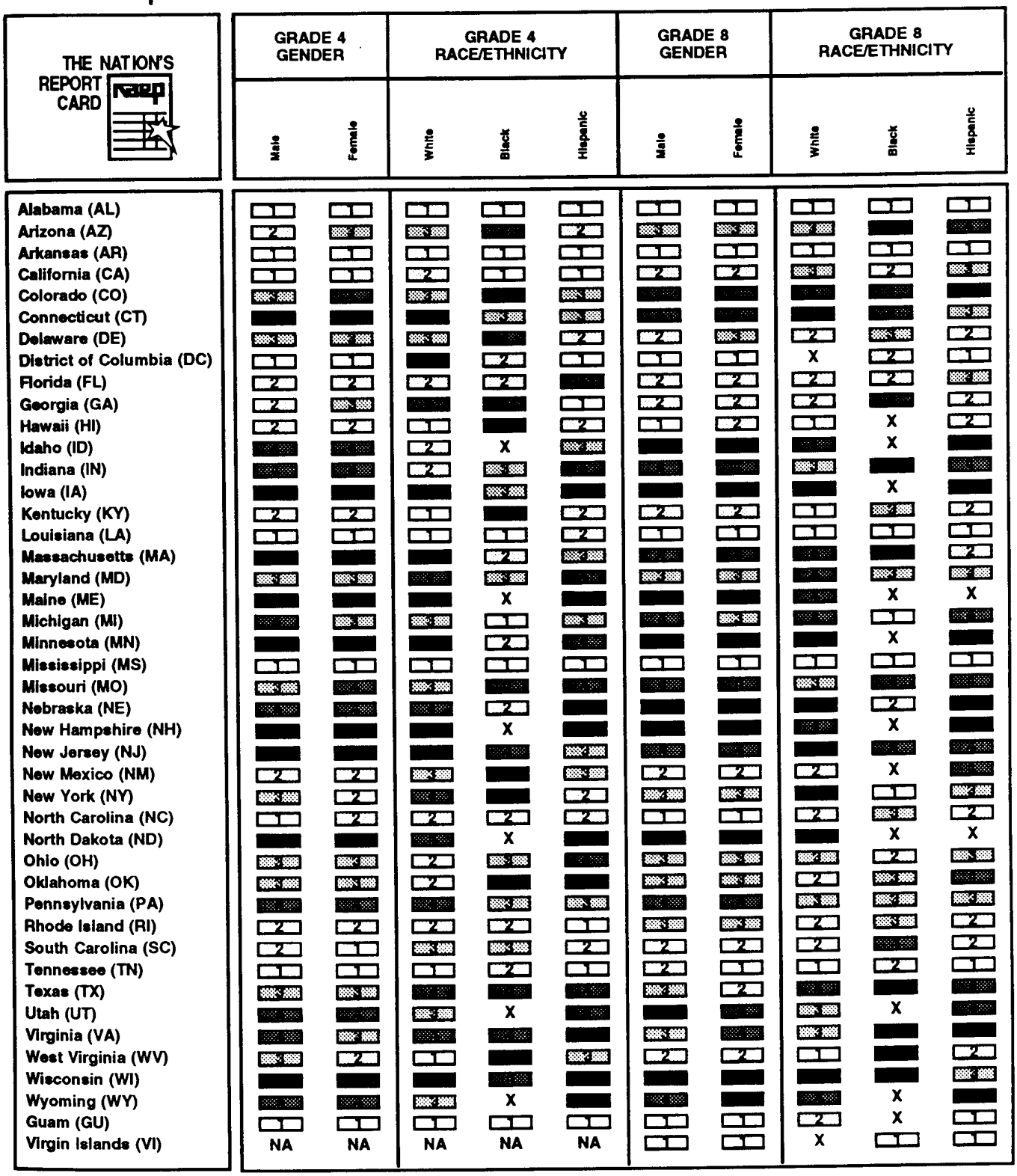
THE NATION'S REPORT CARD



The center *darkest* box indicates a simultaneous confidence interval around the average mathematics proficiency for the state based on the Bonferroni procedure for multiple comparisons. Center boxes that do not overlap indicate significant differences between states in average mathematics proficiency. The *darker shaded* boxes indicate the ranges between the 25th and 75th percentiles of the mathematics proficiency distribution, and the *lighter shaded* boxes the ranges between the 5th to 25th and the 75th to 95th percentiles of the distribution.

FIGURE 6

Average Mathematics Proficiency by Gender and Race/Ethnicity for Five Performance Bands (Quintiles) 1992 Grades 4 and 8



Quintiles of Performance

bottom 20 percent — top 20 percent

1st 2nd 3rd 4th 5th

lower ← proficiency → higher

States categorized in the bottom 20 percent of performance have average mathematics proficiencies in the lowest fifth of the average mathematics proficiency distribution of all states and are indicated by the number 1 (first quintile). States with average proficiencies in the top 20 percent of the distribution are indicated by the number 5 (fifth quintile). The numbers 2, 3, and 4 indicate states with average proficiencies in the second, third, and fourth fifths of the distribution.

X Sample size too small (fewer than 62 students) to permit reliable reporting of performance bands (quintiles).

NA Grade 4 data for the Virgin Islands are not available.

Appendix B

Participants in the Interview Study

Aronson, Lorraine	Deputy Commissioner	Connecticut Department of Education
Benning, Victoria	Education Reporter	Boston Globe
Brigham, Fred	Executive Assistant to the President	The National Catholic Educational Association
Buckley, Cecelia	Administrator of Professional Services	Hampshire Collaborative
Burkhart, Diana	Legislative Attorney, State Legislature	Louisiana
Casserly, Michael	Executive Director	Council of Great City Schools
Chester, Mitchell	Head of Bureau of Curriculum & Instructional Programs	Connecticut Department of Education
Chrostowski, Steve	Researcher	Massachusetts Department of Education
Cohen, Muriel	Education Reporter, Emeritus	Boston Globe
Collins, Angelo	Director	National Committee on Science Education Standards and Assessment
Contois, Donna	State Board of Elementary and Secondary Education	Louisiana
Cooper, Susan	Reading Tutor	Boston Public School
Costello, Karen	Reading Language Arts Consultant	Connecticut Department of Education
Fitzgibbons, Teresa	Human Service Planner, Assessment	Massachusetts Department of Education
Fowler, Mari Ann	Assistant Superintendent of Research and Development	Louisiana Department of Education
Gaudet, Robert	Founder and President	Boston Charter School
Gibbons, Charles	Director	Boston Plan For Excellence
Gregg, Daniel	Social Studies Consultant	Connecticut Department of Education
Halla, Marilyn	Director of Professional Programs	National Council of Teachers of Mathematics
Harvey, Bryan	Supervisor of Campus Assessment Programs	University of Massachusetts

Janiak, Chet	Attorney	Burns & Levinson
Johnson, Susan	Administrator II, Bureau of Pupil Accountability	Louisiana Department of Education
Keefe, James	Director of Research	National Association of Secondary School Principals
Knapp, David	Former University President	University of Massachusetts
Kraft, Betty	Director of Effective Schools	Louisiana Department of Education
Lang, Mic	Assessment Director	Louisiana Department of Education
Leinwand, Stephen	Mathematics Consultant	Connecticut Department of Education
MacCray, Joyce	Director	Council for American Private Education
Miller, Bill	Director of Goals 2000	Louisiana Department of Education
Miyares, Beverly	Supervisor of Research Activities	Massachusetts Department of Education
Moran, Molly	Legislative Assistant	U.S. House of Representatives
Muri, Mari	Mathematics Consultant	Connecticut Department of Education
Murphy, Thomas	Assistant to the Commissioner for Public Relations	Connecticut Department of Education
Natale, Barbara	Consultant (Portfolio Field Trial)	Massachusetts Department of Education
Nolt, Kristin	Legislative Assistant	U.S. House of Representatives
Norton, Scott	Manager, Bureau of Pupil Accountability	Louisiana Department of Education
Park, Hae Seong	Bureau of Pupil Accountability	Louisiana Department of Education
Peat, Stafford	Human Service Planner	Massachusetts Department of Education
Perry, Susan	Consultant	Massachusetts Department of Education
Pruett, Claudia	Administrator II, Bureau of Pupil Accountability	Louisiana Department of Education
Riffel, Rodney	Deals with assessment policy	National Education Association
Rivera, Charlene	Director	Evaluation Assistance Center
Rosenberg, Stan	Senator	Massachusetts

Salus, Richard	Educational Specialist	Massachusetts Department of Education
Sarrat, Marie	Coordinator of LEAP Remediation	Jefferson Parish, Louisiana
Sayer, Gus	Superintendent	Amherst Public Schools
Schuman, Joan	Chief Executive Director	Hampshire Collaborative
Schindler, Jon	Attorney	Klieman, Lyons, Schindler, Gross, & Pabian
Scoffield, Heather	Administrative Assistant	National Council for Geographic Standards
Seidel, Cindy	Superintendent	South Hadley Public Schools
Servat, Yvette	Assistant Director of Secondary Education	Louisiana Department of Education
Sternberg, Betty	Associate Commissioner	Connecticut Department of Education
Story, Ellen	State Representative	Massachusetts
Sumrall, Lois Ann	President, State Testing Commission (and School Principal)	Louisiana
Thomas, Brenda	Human Service Planner, Legislation & Assessment	Massachusetts Department of Education
Tucker, Charlene	Coordinator of Program Evaluation Unit	Connecticut Department of Education
Welburn, Brenda	Executive Director	National Association of State Boards of Education
Weller, Karen	Instruction and Curriculum Specialist	Massachusetts Department of Education

Appendix C

Interview Protocol

Interviewer: _____ Date: _____

Format for the Interviews

Opening Remarks

Begin the interview with some introductory remarks, like those below:

Introduce yourself as working at the University of Massachusetts with Professor Ronald Hambleton on a project for the United States Department of Education. Our project is intended to determine the extent to which educational policy makers and media personnel understand the contents of executive summary reports being produced by the Federal Government (i.e., the Department of Education) to communicate national, regional, and state test results from the National Assessment of Educational Progress (NAEP). (NAEP is a national testing program run by the Federal Government through ETS and Westat. National assessments are conducted every two years with several subjects included in the testing program each time. Reading and mathematics are the most frequently assessed subjects. Only students in Grades 4, 8, and 12 are tested. NAEP is intended to provide accurate information about the status of achievement on important outcomes of schooling, and to provide a basis for monitoring change over time.)

The U.S. Department of Education is concerned that these important educational reports may not be known to policy makers and the media and/or they may not be completely understandable. Problems could be due to the lack of knowledge and experience of the persons reading the reports or due to faults in reporting, or both problems could be present. The results of our interviews should be informative for the Government because they will address the extent of use and understandability of the executive summaries, and suggest ways for improving the reports, if problems are found.

In summary, be sure in your opening remarks to address:

Who we are.

The purposes of our study.

The reason the study is important for American education.

Also, thank participants for their valuable time and interest.

Mention next that our task in the interview is to look through several sections of the 1992 report of the Grade 4, 8, and 12 national and state test results in mathematics. Mention that we will ask some questions, and, along the way, interviewees will provide their thoughts on the format of the report and the results themselves.

Mention that the report consists of 6 sections: (1) major finding of the NAEP study in mathematics, (2) scope of the 1992 national assessment in mathematics, (3) achievement levels (or reporting of NAEP results by achievement levels or what are called proficiency categories), (4) state test results, (5) demographic subpopulation results (e.g., sex and race breakdowns), and (6) results bearing on specific mathematics skills.

Because of time limitations, you can mention we will only be looking at the first and third sections of the report, and one of the sections, four, five, or six.

Background Information

Interviewer: We need to obtain some background information from you because we need to be able to clearly describe participants in the study. We will not use your individual answers anywhere. We are interested only in a summary of the group information we collect but we need names, addresses, and telephone numbers in case some follow-ups to the interview are necessary. Also, we are prepared to mail you a copy of the final report in a couple of months if you are interested in having one. The final report will be completed by the end of October.

(Note: To save time, complete whatever information you can before the interview begins.)

Background Questions

1. Name: _____

2. Address: _____

   _____

   _____

   _____

3. Telephone Number: _____

4. Race (circle one): Black White Hispanic Asian Other

5. Sex (circle one): Male Female

6. Job Description:

7. Work Experience in Education:

   Training in the field of education?

   Work in the field of education?

8. What is your level of interest in national and state student achievement results? (circle one)

   Answer: High Medium Low

9. What is your experience and/or knowledge about educational tests and statistics (e.g., college courses? other training?)? (circle one)

   Answer: None One course More than one course

10. Do you have any knowledge of the National Assessment of Educational Progress, sometimes referred to as NAEP? (circle one)

    Answer: Yes No Unsure

    (**Note:** Skip question 11 if the interviewee answers "No" to question 10.)

11. Have you ever read any NAEP publications in the past? (circle one)

    Answer: Yes No Unsure

    Have you ever seen reports in the newspapers describing NAEP results? (circle one)

    Answer: Yes No Unsure

12. Are you interested in receiving a copy of our final research report when it becomes available at the end of October of this year? (circle one)

    Answer: Yes No

Then say: We are now ready to move to the first section of the NAEP executive summary report of the 1992 mathematics results.

Hand the interviewee a copy of the report, and draw attention to the six sections. (Give the interviewee a chance to flip through the report.) You could mention that similar NAEP reports have appeared in the last couple of years in reading, science, writing and several other subject areas. These reports go back as about 1969. History and geography are being assessed this year.

You could make the interviewee more relaxed by saying that this report is only about mathematics results but that the interviewee should not be concerned if s/he knows little about school mathematics. That's not the focus of the questions and discussion.

Major Findings (Section 1)

Interviewer: Turn to page 1. I would like you to take just a few minutes and read page 1 and on to the middle of page 2, and then we will discuss the material. (Pause, until they finish reading. Perhaps 2 minutes will be sufficient. Be sure the interviewee stops at the middle of page 2.)

13. Please look at the first bullet on page 1. What is being said in the report about mathematics achievement at the national level? (Circle the points or underline the points below that the interviewee identifies.)

(1) at the national level, average mathematics performance improved significantly between 1990 and 1992.

(2) this improvement occurred at all three grades (Grades 4, 8, and 12) and in all types of schools (i.e., public and private).

(Note: If the interviewee makes some incorrect statements, note them below. (Stop them if they go on to describe state results. At this point, state results are not of interest.)

14. In the first bullet we just looked at on page 1 there is a reference to statistically significant increases. (Show them these words.) What do you think these words mean in everyday language? or at least, what do these words mean to you? (circle one)

(Ans. This means that the size of the increase is not just luck or chance. It is large enough that readers can be confident that the difference is almost certainly true. When results are statistically significant, it means we should treat them as if they were true. There is only a small chance that statistically significant results are not true.)

Answer: Correct Incorrect

If the interviewee provides an incorrect answer, write his/her answer below:

15. In the second bullet, there is a quote, "just over 60% of the students in Grades 4, 8, and 12 were estimated to be at or above the Basic level." What do you think this means?

(Ans. The key point here is that the interviewees realize that the 60% applies to the sum of Basic, Proficient, and Advanced students not just Basic students.)

Answer: Correct Incorrect

(If the interviewee answers "40% are below basic," prompt by asking, "And where are the remaining 60%?")

If the interviewee provides an incorrect answer, write his/her answer below:

16. In the third bullet (top of page 2), there is a reference to "considerable variation in performance." What do you think the meaning of this expression is? (circle one)

(Here, we are not looking for a lot of detail. We simply want to know if the interviewee knows that the percent of kids being labeled "Basic," as well as "Proficient" and "Advanced," varies substantially across the states.)

Answer: Correct Incorrect

If the interviewee provides an incorrect answer, write his/her answer below:

Then say, OK let's move on now.

If you have not exceeded the time limit (10 minutes have been allocated from the beginning of the interview to reach this point) you could say that the remainder of this first section highlights the main results of the NAEP Assessment reported by states, by various demographic variables such as race and sex, and looks at some of the findings related to the mathematics curriculum.

Scope of NAEP's 1992 Mathematics Assessment (Section 2)

Interviewer: The next section (pp. 4 to 5), section 2, provides a few details about the size of the national sample of participating students (it is very big—over 250,000), the involvement of the National Council of Teachers of Mathematics (NCTM) (this is very important because the NCTM is the most important mathematics education organization in the country), use of multiple item formats (desirable in an assessment in the l990s because of the shift away from multiple-choice items), and identification of participating states in the trial state assessment portion of the project—1990, Grade 8 only; 1992, Grades 4 and 8 only.

(Here is an important point. You might use this information if you are asked questions about the use of state data in the NAEP Assessment. Students from every state at Grades 4, 8, and 12 participated in the national assessment. These data are used in reporting national results and other important breakdowns. In 1990, 37 of the states, at Grade 8 only, committed to the Trial State Assessment. These states gave tests to extra students and this made it possible in 1990 to report stable mathematics results at the state level too. This was the first time that NAEP results were ever reported at the state level. In 1992 these same states gave tests to extra students so that now it was possible to measure not only 1992 mathematics results in these states but also to

measure growth between 1990 and 1992. Unfortunately, this trial state assessment, as it is called, only involved Grade 8 students and was limited to results in mathematics.

Two other changes took place in 1992 and both are useful for the future. Six new states joined the trial state assessment—including Massachusetts, Mississippi, Missouri, South Carolina, Tennessee, and Utah. All 43 states could obtain state reports for their 1992 mathematics performance, and these state (all 43) are in a position to monitor growth or change in the future. Also, Grade 4 testing was added to the trial state assessment. Grade 4 state results were available in 1992, and the baseline information is available for measuring growth in the future.)

Now let's move on

Achievement Levels (Section 3)

Interviewer: Please take up to 10 minutes and look through pages 6 to 9. (Be sure they don't go past page 9.) In this section the mathematics test results are reported for the various achievement groups: Below Basic, Basic, Proficient, and Advanced.

(Pause to allow the interviewee time to read these four pages.) Then begin the questioning. (Allow about 15 minutes for discussion of the questions below.)

18. Were the definitions of basic, proficient, and advanced students at the top of page 6 clear enough for you to meaningfully read this section of the report? (circle one)

    Answer: Yes No Unsure

19. What changes, if any, would you like to see in these definitions?

    Write suggestions below:

Let's turn now to Table 1.

20. Do you happen to know what the 18% in line 1 means? (circle one)

(Ans. 18% of the Grade 4 students in 1992 were either proficient or advanced.)

Answer: Correct Incorrect

If the interviewee was incorrect, what did he/she think?

21. How about the 1% in line 2? What is the meaning? (circle one)

(Ans. 1% of the students in 1990 at Grade 4 were advanced.)

Answer: Correct Incorrect

If the interviewee was incorrect, what did he/she think?

(Ask the question below only if you have had to explain the interpretation of the numbers in Table 1. It is important to be sure that the interviewee can read this table. Give the interviewee a second chance to show he/she understands the numbers in Table 1.)

Here is a variation. In line 1, what does the 61% mean? (circle one)

(Ans. 61% of the Grade 4 students in 1992 were performing at the Basic, Proficient, or Advanced levels.)

Answer: Correct Incorrect

22. As you look at Table 1, do you see any statistical indicators of growth between 1990 and 1992? (circle the correct answers that they give)

(1) Average proficiency is higher at each grade.

(2) Percentage below basic is less in 1992 than in 1990.

(3) Percent at or above Basic, Proficient, and Advanced levels is higher in 1992 than in 1990.

Write any other correct information they give below:

23. Let's consider next the numbers in brackets. These are called standard errors. Could you figure out from the table (see the footnotes) what they are? (circle one)

Answer: Yes No

(If asked by an interviewee for the meaning, you could say that these standard errors provide an indication of the stability of the numbers to which the standard errors are linked in the Table. For example, consider 218(0.7) in the first line. The correct interpretation is that there is a 95% chance that the true mean proficiency at Grade 4 is between 216.6 and 219.4.)

24. Consider the 61% figure in line 1 of Table 1. The standard error is 1.0. How would you use this standard error? (circle one)

(Ans. If the whole population rather than a sample were used, the true population figure would be between about 59% and 63% or 61% + 2 SEs.)

Answer: Yes No Unsure

Record any errors the interviewee makes below:

25. What do you think the ">" sign means in the table? (circle one)

(Ans. That the number to the left of the sign is significantly greater than the number which follows to the right. Disregard the standard errors. Thus 218 is significantly greater than the 213.)

Answer Correct Incorrect

26. What do you think the "<" sign mean in the table? (circle one)

(Ans. That the number to the left of the sign is significantly less than the number which follows to the right. Disregard the standard error.)

Answer Correct Incorrect

(If the interviewee answers the questions 25 and 26 correctly about the signs, then ask the interviewee to answer question 27.)

27. Go to the table and use these symbols correctly. (Here, we just want interviewees to pick any place in the table where the signs appear, and interpret them correctly. For example, they might say that the Grade 4 mathematics proficiency average of 218 in 1992 is significantly greater than the Grade 4 mathematics proficiency average of 213 in 1990. Or they might say that the 39% of Grade 4 students below basic in 1992 is significantly less than the 46% below basic in 1990.) (circle one)

Answer: Correct Incorrect Did not attempt

28. What is your overall impression of the presentation of information in Table 1? (circle one)

Answer: Clear Needs work Unreadable

29. Do you have any suggestions for improving the communication of information in the table?

Answer:

30. Do you prefer graphs or tables when you are trying to make sense of statistical information? (circle one)

Answer: Graphs Tables No preference Neither (put the information in words)

(Be prepared for interviewees to be confused about Table 1. Expect them to be confused about the entries in the table. For example, they will want to say that 18% of the students in line 1 are proficient rather than the correct statement which is that 18% of the students are proficient or above. Correct this impression if they make the mistake so that they have a fighting chance with the rest of the questions in the interview.)

31. What is your impression of mathematics proficiency based upon your reading of Table 1?

(Ans. Many of the numbers suggest major problems: high numbers in the below basic category, too many students in the basic category, too few students in the advanced category, etc.)

Answer:

Now let's look at Table 2. You may want to check the last sentence on page 6 which mentions these cutpoints.

32. Is the meaning of the numbers in Table 2 clear to you? (circle one)

(Ans. These are the points on the NAEP proficiency scale at each grade level which are used to sort students in the four proficiency categories.)

Answer: Clear Not clear

Record any errors the interviewee makes below:

Follow-up question: What is the meaning of the number 248?

(Ans. It is the score needed to be judged as proficient at Grade 4.)

Answer:

Let's look now at Table 3 which is also on page 7. Here we see the performance of students in the best schools in the country compared to the poorest schools in the country as judged by NAEP results. These performances are reported for each grade separately.

33. If we just focus on the average proficiency scores (column 4) in 1990 and 1992, what is happening? (circle the points identified by the interviewee)

(1) The best schools at Grade 4 and at Grade 8 have shown real improvement between 1990 and 1992. The performances are significantly higher.

(2) The poorest schools have shown much smaller gains between 1990 and 1992 than the best schools, or at least the changes between the two years are considerably less. This result is less clear at Grade 12. In fact, except at Grade 12, performance gains between 1990 and 1992 are nonsignificant.

(Note: the comparison is between Grade 4 students in the best schools in 1990 and the Grade 4 students in the best schools in 1992. Then the same comparison is made for the poorest schools. The analysis is repeated for Grade 8.)

Follow-up question: How does mathematics performance compare in the best high schools and worst high schools in the U.S. in 1992? (circle one)

(Ans. The differences are huge! For example, in 1992 at Grade 12, 82% of high school students in the high performing schools are basic and above. In the lower performing schools, only 40% of students are basic and above. Perhaps the common error is to compare 1990 with 1992 results rather than 1992 top with 1992 bottom schools.)

Answer: Huge Sizable Small No difference

Let's go now to Table 4 on page 9 and look up [Massachusetts, Kentucky, Connecticut, etc. The numbers given as answers below are for Massachusetts.].

(Note: Substitute for Massachusetts whatever state you are in and change the answers below accordingly. If you don't have time to check for the correct answers, simply write down what the interviewee says and score the answers after the interview.)

34. What Is the average proficiency score in [Massachusetts]? (circle one)

(Ans. 226)

Answer: Correct Incorrect

35. How does [Massachusetts] compare to other states? (circle one)

(Ans. [Massachusetts] is one of the higher performing states at Grade 4. The interest here is whether interviewees know enough to look up and down the column with the average proficiency scores. Actually, any of the columns would provide similar information for comparing states. For example, an interviewee could go to the column "Percentage of students at or above Proficient" and use that column to rank the states. By that column, 6 or 7 states would be ahead of Massachusetts.)

Answer: Correct Incorrect

Did the interviewee mention the importance of the standard errors in comparing states? (circle one)

Answer: Yes No

Did the interviewee consider the use of the regional or national information at the top of the table? (circle one)

Answer: Yes No

36. What percent of students in [Massachusetts] are performing Below Basic? (circle one)

(Ans. 30%)

Answer: Correct Incorrect

37. What percent of students in [Massachusetts] are <u>Proficient</u>? (circle one)

(Ans. This is a hard question. 24% are Proficient or above, 3% are Advanced. Therefore, by subtraction, it can be determined that about 21% of the students are in the Proficient category.)

Answer: Correct Incorrect

(At this point the interviewer will need to make a decision. Select either Section 4, 5, or 6 for discussion. Section 6 Is probably best left to interviewees who might have some curriculum interests such as educators.)

Overall Mathematics Performance for the State (Section 4)

(For this next batch of questions, some interviewees may not have any idea how to read the material. If that's the case, you may want to show them how to read this table before proceeding.)

Interviewer: In the little time remaining, I want you to look at the data reported for states. (If Massachusetts, focus on these results. If Connecticut, focus on these results, etc.) This next section of the report allows for the comparison of states with each other.

(Again, choose a state that the interviewee might be interested in.)

Please read the first two paragraphs on page 11 and then look at Figures 1 and 2 on pages 12 and 13.

(Allow 5 minutes or so, more if you have the time and it is needed.)

Then say: If you have read this material, I would like to ask a couple of questions.

38. How many states did significantly better than [Massachusetts]? (circle one)

    (Ans. None)

    Answer: Correct Incorrect

39. How many states did [Massachusetts] outperform significantly? (circle one)

    (Ans. About 24.)

    Answer: Correct Incorrect

40. How do you think [Massachusetts] is doing in these state-to-state comparison results?

    (Ans. Probably the best answer in [Massachusetts] is something like "better than many other states, but the results from an absolute perspective are disappointing. Too many students are Below Basic and not enough students are Proficient.")

    Answer:

Now let's turn to Figure 2. Figure 2 shows the ranking of the states.

41. How did [Massachusetts] rank in Grade 4 mathematics? (circle one)

(Ans. about 8th or 9th)

Answer: Correct Incorrect

42. What do the black bands in Figure 2 represent? (circle one)

(Ans. The mean proficiency scores for the states with confidence bands around the means indicating the instability due to sample sizes.)

Answer: Correct Incorrect

43. Using (say) the 25th percentile point, you could also rank the states. Would the ranking be the same as using the means or average proficiency?

(Ans. Definitely not. The jagged line when you look at the 25th percentile shows clearly that the states would be ranked differently.)

Answer: Similar but definitely not the same

Identical

No idea

44. Why might a policy maker be interested in ranking states based upon 25th percentile, or the other percentiles available in the table? (circle one)

(Ans. Such information gives a clue about how lower performing students are being handled educationally in these states. The presence of special programs, individualized efforts, etc. may be a factor in raising the 25th percentile-like students.)

Answer: Correct Incorrect

Write the interviewee' s response below:

45. Do you have an opinion about the clarity of Figures 1 and 2? (circle one)

Answer: Clear Somewhat clear Confusing Very confusing

Now skip to closing remarks.

Performance for Demographic Subpopulations (Section 5)

Interviewer: This next section on pages 17 to 21 provides information about the performance of various important subgroups. We have time to look at only one or two. Please turn to page 18 and Table 5. I would like you to look at the comparisons of mathematics performance by region of the country near the bottom of the table.

46. First of all, at the Grade 12 level, in which region of the country is the highest mathematics proficiency? (circle one)

(Ans. Central, very closely followed by the Northeast)

Answer: Correct Incorrect

47. Again, at the Grade 12 level, and comparing 1990 to 1992 math performance, which region of the country showed a significant increase in performance? (circle one)

(Ans. the Southeast)

Please read now the last three paragraphs on page 17, which concern Figure 6, and then turn to page 21 and read Figure 6. (Allow interviewees two or three minutes to read.)

48. What do you think is the purpose of Figure 6? (circle one)

(Ans. This figure provides a basis for comparing states with respect to a number of demographic variables.)

Answer: Correct Incorrect No idea

49. What is the interpretation of the numbers in the boxes? (circle one)

(Ans. These numbers tell the quintiles the state is in. High numbers mean the state is doing relatively well and low numbers mean the opposite.)

Answer: Correct Incorrect No idea

50. Is the presentation of information in Figure 6 clear to you? (circle one)

Answer: Yes No Unsure

What Student Know and Can Do in Mathematics (Section 6)

Interviewer: At this point I would like you to take a few minutes and read pages 22 and 23. These pages provide an alternative way to interpret national and state performance. (Pause and allow the interviewee 3 to 5 minutes to read the material.)

51. What do you think is the meaning of the anchor levels?

(Ans. At the anchor points, readers can get an idea about what students know and can do. There are substantial differences among students performing at the four anchor levels.)

Answer:

52. What do you think are the differences between anchor levels and achievement levels?

(Ans. Achievement levels are "shoulds" or expectations; at the anchor levels readers have a good idea about what students can do.)

Answer:

53. Did you find the descriptions of the anchor levels in the table helpful? (circle one)

Answer: Yes No Unsure

54. What do you think information in Table 7 says about the performance of Grade 12 students in the area of reasoning and problem solving involving geometric relationships, algebra, and functions? (circle one)

(Ans. These students are not doing well. Only 6% of students in 1992 were at this level or above.)

Answer: Correct Incorrect

55. In Table 8, we have some state data reported in terms of anchor levels. What percent of Grade 4 students in [Massachusetts] were at a score of 200 or above? (circle one)

(An. 80%)

Answer: Correct Incorrect

56. How does this number [80%] compare to the Nation and the Northeast? (circle one)

(Ans. [Massachusetts] exceeded the national percent by about 9% and the northeast by about 5%.)

Answer: Correct Incorrect

57. What is the significance of the fact that 0% of the Grade 4 students in [Massachusetts] were at a score of 300 or more? (circle one)

(Ans. There is no significance. This is a totally unrealistic target for Grade 4 students. For example, the advanced cut-off score is only 280. Also, no one in the nation exceeded this value. Look at the content expectations, too. This is not Grade 4 work.)

Answer: Correct Incorrect

Closing Remarks

Interviewer: Let me move now to a few final questions.

58. I am going to read a list of changes which have been suggested for improving the Executive Report. Please answer "Yes" if you like the change and "No" if you don't.

a. An introduction describing the purposes of NAEP	Yes	No
b. More use of bullets, boxes, color, checklists, etc., to highlight main points	Yes	No
c. More interpretative information	Yes	No
d. Simpler tables and graphs	Yes	No
e. More complex charts showing inter-relationships among (say) state, race, sex, etc.	Yes	No
f. Executive Summary probably needs to be longer to address all of the important information that is available.	Yes	No

59. Do you have any final thoughts to make about these reports, either format comments or content comments?

Format

Substantive Points

60. How many minutes might you normally expect to spend reading reports like this one? _____ Minutes

That's all we have time for now. You may keep this report if it is of interest to you.

Allocation of Time

Opening Remarks and Collection of Demographic Information	5 min.
Major Findings (Section 1) (2 to 4 minutes reading, 5 minutes on questions)	10 min.
Scope (Section 2) (just mention what it's about)	1 min.
Achievement Levels (Section 3) (pages 6 to 9) (up to 10 minutes reading and 15 minutes on questions)	25 min.
State Performance (Section 4) (pages 11 to 13) (up to 5 minutes reading and 5 minutes on questions)	10 min.
Demographic Subpopulations (Section 5) (pages 17 to 21)	10 min.
What Students Know and Can Do in Math (Section 6) (pages 22 to 28)	10 min.