NATIONAL CENTER FOR EDUCATION STATISTICS

Analytic Report **January 1998**

NAEP'S CONSTITUENTS: WHAT DO THEY WANT?

Report of the National Assessment of Educational Progress Constituents' Survey and Focus Groups

U.S. Department of Education
Office of Education Research and Improvement **NCES 98-521**

THE NATION'S REPORT CARD
naep

What is the Nation's Report Card?

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

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

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

NATIONAL CENTER FOR EDUCATION STATISTICS

NAEP'S CONSTITUENTS: WHAT DO THEY WANT?

Report of the National Assessment of Educational Progress Constituents' Survey and Focus Groups

Roger Levine
Amy Rathbun
Ramsay Selden
Andrew Davis

January 1998

Office of Educational Research and Improvement
U.S. Department of Education

Prepared by the Education Statistics Services Institute under a contract with
the National Center for Education Statistics

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

Office of Educational Research and Improvement
Ricky T. Takai
Acting Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

Education Assessment Group
Gary W. Phillips
Associate Commissioner

January 1998

SUGGESTED CITATION
Levine, R., Rathbun, A., Selden, R., and Davis, A. *NAEP's Constituents: What Do They Want? Report of the NAEP Constituents' Survey and Focus Groups.* Washington, DC: National Center for Education Statistics, 1998.

FOR MORE INFORMATION

Contact:
Arnold A. Goldstein
202-219-1741

For ordering information on this report, write:

National Library of Education
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, DC 20208-5641

or call 1-800-424-1616 (in the Washington, DC, metropolitan area call 202-219-1651).

This report also is available on the World Wide Web: http://nces.ed.gov/naep.

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

Foreword

The National Assessment of Educational Progress (NAEP) has many important constituents: state and local education managers and policy-makers, researchers, the public, practitioners. But, how do these various groups feel about the current features of NAEP and the different directions NAEP could go in the future? Unless we canvas them directly, we have no way of knowing what these important consumers of NCES information want and need.

The NAEP Constituents' Survey was done in the first part of 1997 by the Education Statistics Services Institute as part of an extensive effort to redesign NAEP by the National Center for Education Statistics and the National Assessment Governing Board. Through a comprehensive survey of managers and policy-makers at the local, state, and national levels and through focus groups of other users, project staff asked these groups what directions NAEP should take. What kinds of background questions should it include? How should results be released? How useful is subscale information within the various subjects? Is it important to link NAEP to the results of international studies?

The findings of this effort already have been extremely useful to NCES in guiding future plans for NAEP. We hope you will find them as useful and interesting.

Pascal D. Forgione, Jr., Ph.D.
Commissioner

Acknowledgments

NAEP's Constituents: What Do They Want? Report of the NAEP Constituents' Survey and Focus Groups was authored under a contract with the Education Statistics Services Institute (ESSI), which provides direct support to the National Center for Education Statistics (NCES). The authors, Roger Levine, Amy Rathbun, Ramsay Selden, and Andrew Davis of ESSI/American Institutes for Research (AIR), worked as a team under the general direction of Sharif Shakrani, current National Assessment Governing Board (NAGB) staff member and former project officer at NCES, and Arnold Goldstein, current project officer at NCES.

The authors wish to thank all of those who contributed to the production of this report. Special thanks go to: NCES and NAGB for their evaluation and feedback on draft versions of the NAEP Constituents' Survey; JoAnn Blue, Sarah Benson, Sara Davison, John Konstant, and Freddy Mendez of ESSI, who contacted constituents by telephone to increase survey response rates and assisted in careful reviews of the report; Iby Heller and Kari Wolman of AIR, who also assisted with the follow up phone calling and helped to prepare the summary of responses to open-ended comments; and Fran Stancavage of AIR, who provided valuable input into the development of the survey items and focus group protocol.

We would also like to thank several individuals who served as reviewers and provided many insightful and helpful comments. The reviewers were: Peggy Carr, Patricia Dabbs, Arnold Goldstein, Steven Gorman, Andrew Kolstad, and Kelly Weddel of NCES; Ray Fields of NAGB; Doris Redfield, a consultant; and Jeff Nelhaus of the Massachusetts Department of Education.

Lastly, thanks are due to the chief state school officers, state education agency assessment directors, state education agency curriculum directors, state board of education chairs, Governor's education policy advisors, state legislature education committee staff, superintendents of large urban and suburban school districts, staff of education associations, and education researchers who responded to the survey; and to the principals, teachers, members of the media, representatives of national business organizations which conduct efforts to support and improve schools, and members of the public who participated in the focus groups. Their input to this study was vital to the redesign of NAEP.

Table of Contents

Executive Summary

Purpose

The National Assessment of Educational Progress (NAEP) remains the only accurate and credible indicator of educational performance capable of informing about both national trends and state differences in student achievement. NAEP, which is also known as "The Nation's Report Card," tests fourth, eighth, and twelfth-grade students in reading, writing, mathematics, science, history, geography, civics, the arts, foreign language, and economics. However, NAEP has become less efficient and more complex and costly to administer over the years. As a result, it may be that not enough subjects are tested, and test results may not be reported to the public soon enough after the tests. In addition, an assessment designed to provide only national or state-level measures may not be meeting the needs of states or their school districts. Many people have suggested that the time has come for NAEP to be redesigned, so it can monitor the educational achievement of students in our nation in a more efficient and comprehensive manner and better serve the informational needs of local districts, states, and others -- without sacrificing its quality, accuracy, or reliability.[1]

NAEP serves many different constituencies whose opinions must figure heavily in determining the future directions NAEP should take. However, what these constituent groups want usually is not measured systematically. It is gleaned anecdotally or through processes of consultation or consensus building that are not systematic enough to give accurate indications of the group members' true feelings. Furthermore, the feelings of different groups are not usually compared and weighed in relation to one another. This study measured constituency opinions directly on key features and directions of NAEP.

Method

In order to directly and precisely measure, analyze, and compare the opinions of some key NAEP constituencies, a multi-stage process was undertaken:

(1) Key constituencies and issues were identified through consultation with NCES and NAGB. Groups (constituencies) included Chief State School Officers, State Assessment Directors, State Curriculum Directors, Governors' Education Aides, State Board Chairs, State Legislature Education staff, Superintendents of large urban and suburban districts, and senior staff of national education organizations.

(2) Draft survey instruments to measure opinions about the issues were developed and pilot tested by a team of researchers from the American Institutes for Research (AIR) and the Education Statistics Services Institute (ESSI). After the

[1] National Assessment Governing Board, Policy Statement on Redesigning the National Assessment of Educational Progress, August 2, 1996.

instruments were reviewed and approved by the Office of Management and Budget (OMB), surveys were administered to representatives of the eight key constituencies (groups).

(3) Focus groups were conducted to assess opinions about these issues with representatives of other constituencies. Focus groups provide less precise indicators of how different constituencies feel about NAEP than can be obtained with surveys. Logistical considerations required their use to provide these other constituencies with opportunities for input into the redesign process.

Only issues in the NAEP redesign which were open -- for which policy decisions had not been made -- were investigated in this project. The survey topics and constituents' responses are highlighted in the Results section of this summary.

Surveys

A draft NAEP Constituents' Survey was developed and reviewed by NCES and NAGB. During the development of the survey, three basic principles were followed:

- Survey items had to be closed-ended to encourage a high response rate.
- Additional comments and insights would be allowed through opportunities for respondents to expand upon their responses.
- To allow evaluation of different alternatives, their major benefits and costs or tradeoffs would be described in each item.

After review, pilot testing of the survey was conducted and minor modifications were made based on feedback from participants. Three versions of the survey were prepared to reflect the differences in knowledge and interests of different groups.

After OMB approval, the NAEP Constituents' Survey was mailed to 424 constituents in late February, 1997. All members in each of the groups were surveyed. Within each group, there was typically one representative from each state. Table A identifies the groups surveyed, the number of constituents surveyed, the number of returned surveys from each group, and the final response rates. Individuals in different roles within the states were strongly encouraged to respond from the perspective of their roles, as opposed to deferring to other respondents. Further, states were urged not to attempt to create a "unified" response, so the distinct opinions of each group across states could be measured.

Intensive follow-up procedures were used to increase response rates to the mail survey. Telephone calls to nonrespondents were made by trained AIR/ESSI staff in March and early April. With telephone follow-up, the overall response rate was 83 percent (352 completed surveys).

Survey analyses included summaries of response frequencies for all items and cross tabulations by constituent group. Statistical tests were conducted to ensure that differences in responses across the different types of constituencies were not due to chance. The results of the analyses are summarized in the Results section of this summary.

Table A. Overall Response Rate, by Respondent Group

Respondent Group	Number of potential respondents	Number of surveys received	Final response rate
State Education Agency Assessment Directors	52	51	98%
State Education Agency Curriculum Directors	51	48	94%
Large Suburban School District Superintendents	49	45	92%
Large Urban School District Superintendents	48	41	85%
Chief State School Officers	51	43	84%
Staff of Education Associations	19	16	84%
State Board of Education Chairpersons	52	37	71%
State Legislature Education Committee Staff	51	36	71%
Governors/Education Policy Aides	51	35	69%
Total	**424**	**352**	**83%**

Focus Groups

For some groups of constituents, the NAEP Constituents' Survey would not be feasible to administer. Protocols for focus groups were developed to parallel the structure of the basic survey. Seven focus groups were held to gather the input of:

(1) public high school principals,
(2) private high school principals,
(3) elementary school principals,
(4) members of the general and education press,
(5) the general public,
(6) members of national business organizations which conduct efforts to support and improve schools, and
(7) teachers.

The focus groups were not mixed across constituencies (i.e., elementary school principals were interviewed with other elementary school principals and not with teachers, etc.) In total, 46 individuals from 19 different states and the District of Columbia participated in the focus groups.

To solicit input from the research community, the NAEP Constituents' Survey was sent to all current American Education Research Association (AERA) and National Council on Measurement in Education (NCME) officers and their immediate predecessors. Since their concerns were expected to be very different from those of the constituencies involved with the implementation of NAEP, these results were analyzed separately. They are reported in the Focus Group section.

Results

Several topics were covered in the NAEP Constituents' Survey and the focus groups. The following sections highlight constituents' responses to each issue.

Emphasis of Background Information

Constituents were asked to identify how important they felt four types of background characteristics were for NAEP to study. This input was used to prioritize the importance of 1) school characteristics; 2) student background factors; 3) instructional practices; and 4) topics of current educational relevance.

- Survey respondents felt that instructional practices and student background factors were significantly more important than school characteristics or topics of current educational relevance.

- Focus group participants expressed that school and student background characteristics were most important, followed by instructional practice information. Topics of current educational relevance were not mentioned frequently as being important background variables to measure.

Impact of Background Questions on Release of Results

Respondents were asked to consider the value of background items in light of their impact on when results could be released. Opinions were similar for all constituents.

- Few respondents felt that all background items should be removed from reporting in order to shorten the report production time a great deal.

- Most respondents agreed that some delay in reporting was acceptable, since background variables provided valuable information. Some suggested keeping background questions but reducing the number included.

- Several respondents suggested a multi-step release, consisting of an initial report with no or limited background information followed by a later comprehensive report with full background information.

Impact of Technical Documentation on Release of Results

Constituents were also asked to consider the necessity of technical documentation -- that is, the extensive documentation of the psychometric and other technical characteristics of the assessment -- in NAEP reports in light of its impact on the report release date.

- All constituent groups that were surveyed felt the amount of documentation of the technical (i.e., psychometric) characteristics of the assessment could be reduced so that results could be reported sooner. Some constituents suggested that technical information could be made available on the Web.

- However, the focus group principals and teachers felt that the reports should contain full technical documentation upon release, even if it led to a delay in the reporting date.

- A multi-step approach to reporting was again suggested by some constituents.

Inclusion of a Parent Survey

Respondents were asked to consider the tradeoffs -- that is, the value of the information versus its costs -- of including a parent survey as part of NAEP. Opinions differed on this topic; however, most groups noted common benefits and potential concerns in collecting data from parents.

- Most survey respondents were opposed to the idea of including a parent survey, especially those who would have the most responsibility for implementing the data collection (e.g., State Assessment Directors). The media and business community members attending the focus groups were also opposed to a parent survey.

- State Legislative Education Committee staff, education researchers, and the members of the principals', teachers', and general public focus groups were in favor of a parent survey. However, all noted the logistical problems of data collection (e.g., potentially low response rates, intrusiveness, need for translations for non-English speaking parents).

Schedule for Release of Results

Respondents were asked whether they prefer to have reports released as they become available (the current procedure) or as part of a predetermined schedule. The responses differed somewhat again between the survey and focus group participants.

- Survey respondents (state and local policy makers and administrators) were nearly equally divided between those who preferred a set schedule and those who preferred the current practice. Of those who preferred a set schedule, the majority favored releases at different predetermined times throughout the year, as opposed to releasing all reports within the same week.

- Most focus groups, with the exception of education researchers[2], preferred reports to be released on a set schedule. The set schedule would help them better prepare and educate the public on the purpose and results of NAEP. The media focus group participants would prefer that results are not all reported within one week, since writers would not be allotted adequate space to cover all results within such a short time frame.

State Mandates for NAEP Participation

State representatives were asked to indicate whether their state mandates NAEP participation.

- Most states do not mandate participation in either state or national NAEP, nor do they plan to do so in the future. States that do mandate participation are more likely to do so for the state NAEP than for the national NAEP.

Subjects Assessed at the State Level

State representatives were also asked to indicate how likely it would be that their state would participate in various years of NAEP state assessments.

---

[2] Education researchers preferred to have NAEP results released as soon as they were available.

- Three-quarters of the states thought they would participate in every state NAEP through the year 2010, although uncertainty about participation increased slightly over time.

- Ninety percent of the states expressed interest in participating in state assessments for social sciences and history, and about fifty percent of the states indicated an interest in state assessments for the arts and foreign languages, even though these subjects are not currently scheduled for state-level reporting.

Assessing Subject Areas in Combinations or Individually

Respondents were asked about their preferences for measuring and reporting results for social sciences and history, natural sciences, and reading/language arts; that is, whether they would prefer to have scores in each of the areas reported for each individual (component) subject or reported as a cluster.

- All constituent groups felt strongly that reading and writing should be assessed as individual subjects.

- Survey respondents indicated that social sciences and history and natural sciences should be assessed and reported as a cluster, in contrast to current practice.

- Within each of the focus groups, there were mixed opinions expressed about the assessment of natural sciences and social sciences and history. About half of the participants believed the subjects should be assessed and reported separately; the other half favored a cluster assessment.

- When asked to indicate a preference for integrated or combined assessments, State Assessment and Curriculum Directors tended to prefer use of integrated approach (i.e., subject clusters assessed and reported in relation to one another) over a combined approach (i.e., subject clusters assessed and reported as free-standing areas combined in the same assessment).

Desire for Information on Skill Areas within Subjects

Constituents were asked to indicate how important subscale (skill area and discipline) scores, in addition to overall scores, were for the subjects of mathematics, reading, writing, science, history, and geography at each of the NAEP grade levels (4, 8, and 12). Similar preferences were noted by the survey and focus group respondents.

- For the survey respondents, subscale scores for math were most important, followed by subscale scores for reading, writing, science, history, and geography. Further, subscale scores were viewed as less important for fourth graders.

- Responses from the focus groups were similar; however, several groups felt that subject breakdowns were necessary for all grade levels, with the possible exception of history and geography in fourth grade. They indicated that subscale scores were useful for guiding state curriculum planning and improvement. The groups expressed some concerns about the current labels for breakdown categories in the various subject areas.

Frequency of Data Collections

In the past, NAEP tested students every other year. Congress recently authorized NAEP to collect data every year. Respondents were asked to indicate whether they would prefer a yearly administration, and if so, why they preferred it to a biennial schedule (i.e., administration of NAEP every other year).

- The majority of constituent groups preferred the annual assessment to the current biennial schedule.

- Several respondents expressed concern about the increased burden on schools. Principals in the focus group felt the change to annual testing would not reduce school burden since the classroom setting is disrupted whenever NAEP occurs, regardless of the number of subjects assessed or the number of students tested. Their concern was the number of times their school had to participate in testing rather than the number of students tested at each time.

- Teachers and principals participating in the focus groups preferred that annual assessment be used to assess subjects more frequently, as opposed to assessing more subject areas.

Linking NAEP to International Assessments

Constituents were asked about the value of linking NAEP scores to international assessments, to allow individual state performance to be compared with the performance of other countries. Similar opinions were expressed by all constituent groups.

- All constituent groups participating in the survey and focus groups favored the government supporting this type of research.

- Respondents felt that the government needed to conduct such research in a quality way if comparisons were to be made, indicating that funds should be allocated to meet more than minimal standards of reliability and validity.

Obtaining State-Level Results

State education assessment and curriculum directors were presented with various approaches to obtaining state-level information on NAEP, including the option of a "market basket" approach, which would provide representative sets of assessment exercises, and the current, full state assessment. In the market basket approach, short modules representing the assessment would be offered to states. These modules could be used to obtain state-representative results, to calibrate state assessments to the NAEP scale, or to obtain state-level information more frequently than the NAEP-funded schedule would provide. Respondents were asked about their interest in using the current state NAEP assessment, the market basket assessment, or a combination of both measures.

- Comparable proportions of states indicated that they were definitely interested in the current state assessment and the market basket program (39 and 31 percent, respectively). Another 10 percent were interested in using both measures.

Constituents were also asked how important seven factors (i.e., state costs, school burden, psychometric test properties, ability to compare results with other states, ability to use results for within state comparisons, ability to obtain information on instructional practices and their relationship to student achievement, ability to obtain student-level results) were in evaluating alternative approaches for obtaining state NAEP results.

- State costs and school burden were significantly more important than any other factors, with the exception of psychometric test properties.

- The importance of the seven factors differed significantly among the various constituencies that were surveyed in the states.

States Paying for Some Services

Respondents were asked to assess their state's willingness to pay for three different services: a state level assessment using the current approach; linking NAEP results with the state's regular assessment; and the provision of extra "market basket" assessments for states to use as they desire.

- Overall, states indicated limited willingness to pay for most services. However, about two thirds of the states indicated a willingness to pay for NAEP linkages to state assessments.

Conclusions

Although the constituent groups did not have identical opinions on the issues discussed in the NAEP Constituents' Survey and focus groups, there were many cases in which similar views were expressed. Attitudes which were common to most or all of the respondents include:

- Instructional practices, student characteristics, and school characteristics are important background variables to measure on the NAEP assessment.

- Some of the above background variables need to be included in NAEP reports, even if they delay the release of reports, but the number could be reduced or handled differently to make the process more efficient.

- A multi-step approach could be used in reporting NAEP results to eliminate delays due to the inclusion of background variables and technical documentation. A preliminary report that contained only achievement data could be released, followed by a comprehensive report at a later date.

- Most states do not mandate participation in national or state NAEP, nor do they plan to mandate participation in the future. However, most states indicated that they will participate in all yearly NAEP state assessments through the year 2010. Most states would be very interested in participating in state assessments in social sciences and history, the arts, and foreign languages.

- Reading and writing should be assessed as individual subjects; other subjects could be assessed in clusters.

- Subscale scores (in addition to overall scores) were viewed as most important for math. Subscale scores were viewed as being less important for fourth graders than for students at the higher grade levels. Overall, subscale scores were viewed as important.

- NAEP assessments should be administered annually instead of biennially.

- The government should definitely support efforts to link NAEP to international assessments. It should fund such research to the extent that reports meet more than minimal standards of reliability and validity.

- In the evaluation of alternative assessment approaches for obtaining state NAEP results, state costs, school burden, and psychometric properties are the most important factors in decision making.

- States would be more willing to pay for NAEP linkages to state assessments than they would be to fund extra market basket assessments or the current state NAEP.

NAEP's Constituents: What Do They Want?
Report of the NAEP Constituents' Survey and Focus Groups

1.0 INTRODUCTION

The National Assessment of Educational Progress (NAEP), which is also known as "The Nation's Report Card," tests fourth, eighth, and twelfth-grade students in reading, writing, mathematics, science, history, geography, civics, the arts, foreign language, and economics. (The choice of these subjects was determined by Congress.)

The purposes of the National Assessment have remained the same since NAEP was first administered in 1969:

- to provide data about what American students know and can do and
- to monitor trends in student educational performance over time.

With the 1990 mathematics assessment, the National Assessment began reporting results using student performance levels, called achievement levels, set by the National Assessment Governing Board. The three achievement levels -- basic, proficient, and advanced -- are reported to help the public interpret how well students are performing on the National Assessment. Explaining the reasons *why* educational achievement has or has not changed is not the central purpose of NAEP. It is more like a barometer, measuring levels and changes rather than trying to explain why they occur.

In order to provide states with measures of student performance, an optional state-level NAEP was implemented. If a state chooses to participate, an additional sample of schools in that state are selected and their students tested. With these additional schools, it is possible to produce state-level estimates. State-level results have been reported (for participating states) since 1990. Participation is optional for each state. In 1996, approximately 46 states and jurisdictions participated in the state NAEP program (formerly known as the "Trial State Assessment").

The National Assessment of Educational Progress remains the only accurate and credible indicator of educational performance capable of informing about both national trends and state differences in student achievement. However, as pointed out in the National Assessment Governing Board's (NAGB's) Policy Statement on Redesigning the National Assessment of Educational Progress[3], NAEP has become less efficient and more complex and costly to administer over the years. As a result, not enough subjects can be tested and test results are not reported to the public soon enough after the tests. In

[3] The National Assessment Governing Board's (NAGB's) Policy Statement on Redesigning the National Assessment of Educational Progress was adopted unanimously by NAGB on August 2, 1996.

addition, an assessment designed to provide only national or state-level measures may not be meeting the needs of states or their school districts. Many people have suggested that the time has come for NAEP to be redesigned so that it can monitor the educational achievement of students in our nation in a more efficient and comprehensive manner and better serve the informational needs of local schools, districts, and states -- without sacrificing its quality, accuracy, and reliability. The National Center for Education Statistics (NCES) has expressed a commitment to address the concerns enumerated in NAGB's redesign policy.

NAEP serves many different constituencies, whose opinions must figure heavily in determining the future directions that NAEP should take. Their desires are often put forward by individuals or organizations, arguing for NAEP to go in a particular direction. However, what these constituent groups want usually is not measured systematically. It is gleaned anecdotally or through processes of consultation or consensus building that are not systematic enough to give accurate indications of the group members' true feelings. Furthermore, the feelings of different groups are not usually weighed in relation to one another.

NCES commissioned this study to systematically measure the perceptions of a set of constituencies with respect to the key features and directions of NAEP. In order to directly and precisely measure, analyze, and compare these opinions in some key constituencies, a multi-stage process was undertaken:

(1) Key constituencies and issues were identified through consultation with NAGB and NCES,

(2) Draft survey instruments to measure opinions about these issues were developed and pilot tested by a team of researchers from the American Institutes for Research (AIR) and the Education Statistics Services Institute (ESSI). After the instruments were reviewed and approved by the Office of Management and Budget (OMB), surveys were administered to representatives of eight key constituencies (groups), and

(3) Focus group guides to assess opinions about these issues were developed and administered with representatives of other constituencies. Focus groups provide less precise indicators of how different constituencies feel about NAEP than can be obtained with surveys. Logistical considerations required their use to provide these other constituencies with opportunities for input into the redesign process.

The procedures employed in the performance of these tasks and the information collected are summarized in this report.

1.1 Identification of Constituencies and Issues

An issues-by-group matrix was developed after consultation with NCES. The initial constituencies proposed included:

- Governors and/or their Education Policy Advisors/Aides
- Chief State School Officers
- State Assessment (Test) Directors
- State Curriculum Directors
- State Board Chairs
- Educational Organizations

After discussions, it was decided that the superintendents of the 50 largest urban and 50 largest suburban school districts in the nation should also be surveyed.

Surveys of state-level representatives of teachers and school administrators were considered. However, the logistics involved in the creation of a sampling frame and in the selection of a representative sample lead to the decision to collect the opinions of these constituencies through focus groups conducted at national meetings. Similarly, it was determined that the opinions of businesses and the media would also be collected through focus groups, for reasons of logistics. The opinions of the general public were never intended to be collected through surveys. Instead, their feelings about NAEP would also be assessed through use of focus group methodologies.

Issues to be investigated that were initially proposed included:

- Subjects assessed at the national and state levels
- Schedules for assessing various subjects
- Interest on information on subscales within a subject
- Preferences on frequency of testing
- Attitudes toward major alternatives for providing state-level data
- Preferences for trade-offs on background data versus timeliness of reporting
- Preferences for focus of background information (instructional practice, demographics, etc.)
- Preferences on trade-offs between timeliness and technical documentation

Two of these areas were eliminated: subjects assessed at the state and national levels and schedules for assessing various subjects. The areas were eliminated since decisions about testing schedules had already been made. Only areas of NAEP redesign for which policy decisions had not yet been made would be investigated in this project.

After discussions with NCES and NAGB, the following areas were added to the research agenda:

- Inclusion of a parent survey
- Schedules for releasing results
- Current and future state mandates for participation in NAEP
- Assessing subject areas in combinations (clusters) or individually, and multi-subject scoring
- Linking NAEP to international assessments
- Willingness of states to pay for different types of NAEP services

1.2 NAEP Constituents' Survey

1.2.1 Survey Instrument Development

A draft survey instrument was developed and circulated for internal review within ESSI and AIR. After modifications, the revised draft was reviewed by NAGB and NCES. Underlying the development of this instrument were several basic principles:

- Survey items had to be closed-ended to encourage a high response rate.
- Additional comments and insights would be elicited through the provision of opportunities for respondents to expand upon their responses.
- To allow evaluation of different alternatives, their major benefits and costs would be briefly described as part of each item.

Based on comments and suggestions from both NAGB and NCES, the instrument underwent further revisions, and a version was produced for pilot testing. Pilot testing on both the east and west coast was conducted with five individuals, representing four of the various constituencies to be surveyed. Respondents completed questionnaires and underwent a thorough debriefing. This debriefing was conducted by staff trained in the administration of a special questionnaire debriefing guide that was prepared for this

purpose. As a result of the pilot testing, the questionnaire underwent further minor modifications. Final versions were prepared and submitted and approved by OMB.

Three different survey versions were prepared, reflecting the differences in knowledge and interests of the various constituencies (see table 1-1). Survey versions I and II were identical, except for an item (8B) that was asked only of State Education Agency Assessment and Curriculum Directors (see Appendix A). Survey version III was the shortest, containing only 10 of the 14 items included in version II.

1.2.2 Survey Mailout

On February 25, 1997, surveys asking for comments and suggestions on future directions of the National Assessment of Educational Progress were sent to 424 potential respondents in the 50 states, Washington, D.C., and Guam[4], including:

- Governors/Education Policy Aides
- Chief State School Officers
- State Board of Education Chairpersons
- State Education Agency Assessment Directors
- State Education Agency Curriculum Directors
- State Legislature Education Committee staff
- Education Association staff
- Superintendents of large urban and suburban school districts

Although it was intended that each of the first six listed surveys be sent to each state (and D.C. and Guam), the indicated position was either vacant or missing in some states. Similarly, although it was intended that the largest 50 suburban and the largest 50 urban school districts be surveyed, two of the districts in the sampling frame were also state education agencies. The third case was a district defined to be out-of-frame: Puerto Rico.

[4] Surveys were prepared for the 424 individuals who were believed to be filling these roles. However, surveys were actually mailed to only 412 potential respondents because 10 of these individuals also filled the indicated role for another respondent group and one individual was listed as filling three different roles. These 11 individuals were each sent only one copy of the survey to complete and return. The responses on their completed surveys were then transferred to the duplicate surveys.

Telephone contact records to document each call were created for each survey recipient. The contact logs were color-coded to match the version of the survey that was sent or administered to the respondent. Thus, respondents with green contact logs were administered green (version I) surveys.

Table 1-1. Summary of Surveys Mailed to NAEP's Constituents

Respondent Group	Number of potential respondents	Survey version	Survey color	Listed respondent
State Education Agency Assessment Directors	52	I	green	probable completer
State Education Agency Curriculum Directors	51	I	green	probable completer
Governors/Education Policy Aides	51	II	blue	probable delegator
Chief State School Officers	51	II	blue	probable delegator
State Board of Education Chairpersons	52	II	blue	probable delegator
State Legislature Education Committee Staff	51	II	blue	probable delegator
Staff of Education Associations	19	III	pink	probable delegator
Large Suburban District Superintendents	49	III	pink	probable delegator
Large Urban District Superintendents	48	III	pink	probable delegator
Total	**424**			

Included in the mailout to all potential respondents were a cover letter, a postage-paid return envelope, and the appropriate version of the survey. It was expected that most of the listed respondents would delegate the responsibility for completion of the survey, so the majority of individuals to whom questionnaires were sent were provided the option of completing and returning the survey themselves, or delegating responsibility for completion to another staff member *in their office*. The procedures for delegation were explained in the cover letters accompanying the surveys to all potential respondents except the Assessment and Curriculum directors. These individuals -- the probable delegators -- were also sent a postcard to return with information about their designee if they chose to delegate responsibility for completion of the survey.

Individuals in different roles within the states were strongly encouraged to respond from the perspective of their roles, as opposed to deferring to other respondents. Further, states were urged not to attempt to create a "unified" response. This was

requested so that the distinct opinions of each group could be measured.

1.2.3 Follow-up Operations

In order to increase the response rate to the initial mailing, nonresponding survey recipients or their designees were contacted by telephone. These phone calls commenced on March 11, 1997 and were conducted by AIR/ESSI staff in Palo Alto and Washington, D.C. The calls were intended to:

- confirm that the surveys had been received. If not, duplicate surveys were sent by either fax or Federal Express.
- determine if the survey recipient intended to complete and return the survey, or intended to delegate responsibility for its completion to a staff member.
- answer any questions the survey recipients or designees might have had about the survey or the data collection effort.
- encourage completion and return of the questionnaire. If the respondent was willing, the survey could be completed over the telephone during the call or during a scheduled call-back.

A Paradox-based Management Information System was developed and used for record keeping processes.

Before these phone calls commenced, a training guide was prepared for the follow-up staff to review before attempting to contact the survey recipients and designees, and to use as a guide during their conversations with the nonrespondents. This training guide included:

- an overview of NAEP and the redesign effort,
- a summary of the surveys mailed to the constituents,
- helpful telephone hints and suggestions to keep in mind when making the calls,
- general approaches for phone calls to nonrespondents,
- interview guides for phone calls to nonrespondents,
- procedures for sending additional surveys by fax or Federal Express,
- record keeping and administrative procedures,
- a guide for administration of the survey for phone interviews, and
- problems that might arise and how the callers should deal with them.

The contents of the training guide were reviewed in detail with follow-up staff at separate training sessions that were held in Washington, D.C. and Palo Alto, CA. To supplement the training guide, mock interviews were also employed. These mock interview exercises provided staff with an opportunity to engage in follow-up activities that simulated anticipated "real-life" situations. One mock interview was a hypothetical conversation between a caller and an uncooperative staff member; the other interview was a hypothetical conversation between a caller and a survey recipient who had several complaints about NAEP.

The follow-up phone calls were extremely successful. Without the follow-up phone calls to the nonresponding survey recipients, a high response rate would not have been possible. As of March 11th -- the deadline for completion and return of the survey -- completed surveys had only been received from 62 of the 424 survey recipients (15%). The follow-up phone calls to the nonrespondents commenced the afternoon of the survey deadline, and by March 19th completed surveys had been received from 41% of the survey recipients or their designees. The data from these surveys were analyzed to allow presentation of preliminary findings on March 26, 1997 at the American Educational Research Association (AERA) conference in Chicago.

The follow-up phone calls to the nonrespondents continued until April 15th. Ultimately, completed surveys were obtained from 352 of the 424 survey recipients or their designees -- an 83% response rate (see table 1-2). Postcards were received from 65 of the survey recipients indicating the name, title, and phone number of their survey designees. All but one of the 65 designees completed and returned the survey as instructed. At least two responses were received from each state, and from the majority of states, five or six responses were received.

Table 1-2. Overall Response Rate, by Respondent Group

Respondent Group	Number of potential respondents	Number of surveys received	Final response rate
State Education Agency Assessment Directors	52	51	98%
State Education Agency Curriculum Directors	51	48	94%
Large Suburban School District Superintendents	49	45	92%
Large Urban School District Superintendents	48	41	85%
Chief State School Officers	51	43	84%
Staff of Education Associations	19	16	84%
State Board of Education Chairpersons	52	37	71%
State Legislature Education Committee Staff	51	36	71%
Governors/Education Policy Aides	51	35	69%
Total	**424**	**352**	**83%**

Unfortunately, not all of the constituents to whom surveys were sent could be persuaded either to participate or to designate another staff member in their office to complete and return the survey. Of the 72 potential respondents in our sample who did not return a completed survey, 12 (3%) explicitly refused and explained why they refused to complete the survey. The most common reasons were:

- the person was extremely busy,
- NAEP does not impact their state because they do not participate in NAEP,
- the person thought he or she was not knowledgeable enough about NAEP, or
- the person thought that he or she was not the appropriate person to fill it out.

For example, one governor wrote:

> *"At this time I am unable to answer your survey for several reasons. First, I believe that the appropriate persons to review this survey (the Department and the Board) have already had the opportunity to do so. Also, I do not have adequate information as to the purpose or mission of the National Assessment of Educational Progress (NAEP) at this time. Finally, I do not possess adequate information as to*

> *STATE's current or future involvement with NAEP. Again, I appreciate the opportunity to participate in your survey."*

Similarly, a district superintendent wrote:

> *"As a member of this school district, I have never heard anyone refer to NAEP. We have a state assessment program which tests 3rd through 8th graders, 10th graders, and at the exit level. This state assessment is the obsession of everyone in the district and our only concern in terms of standards. If our kids are OK on (state assessment), why do we need NAEP? It doesn't impact us. Our state has standards, and the comparisons it draws are what gets our attention. National and state to state comparisons are of little interest."*

The remaining 60 (14%) survey recipients from whom we did not receive completed surveys never informed us of their reasons for nonparticipation, even though each one was contacted on several different occasions to encourage them to complete and return the questionnaire.

1.2.4 Data File Preparation

Prior to data analysis, survey data underwent preliminary edits and cleaning. This involved manual review of each instrument to ensure that the respondents:

- checked only one box (when applicable),
- circled only one response (when applicable),
- observed the skip patterns,
- provided responses that appeared reasonable and appropriate based on responses to other questions, and
- returned only one copy of the survey for each office.

After each item, the respondents were also provided space to elaborate on any of their responses. All comments from the respondents were entered into a large Microsoft Word document and were organized by survey item. To ensure confidentiality, all references to individual states (or districts) were removed from the comments. A summary of the comments from the respondents was also prepared for each of the survey items (see Appendix B).

1.2.5 Data Analysis: Response Frequencies

Simple response frequencies were prepared for all items. These frequencies were then cross-tabulated by type of respondent -- that is, the constituency of which the respondent was a member.

Certain items (questions 6, 7, 12, and 13) were summarized using state as the unit of analysis. Since respondents in several different roles were responding from each state, a means was needed for determining the response for the state. For question 6, which asked whether participation in the state and national NAEPs were mandated, the response of the State Assessment Director was used as the state's response, since that individual was felt to be knowledgeable on that question. For items 7, 12, and 13, the following procedures were used to determine appropriately a single response for each state:

(1) Different weights were applied to an individual's survey responses, as a function of the constituency represented. Weights of 2 were applied to responses from the Chief State School Officer and the State Assessment Director, since those individuals were felt to have the greatest amount of information or the most relevant opinion on most matters; weights of 1 were applied to responses from the State Board of Education Chair, the Governor's Education Aide, and State Legislative staff; and a weight of 0.5 was applied to responses from the State Curriculum Director. (These items were not administered to Association staff or Superintendents of large school districts.)

(2) The response with the greatest weighted sum was identified as the state's response.

(3) If two or more responses were tied for the greatest weighted sum, the response was left blank.

It is recognized that this approach is not sensitive to the fact that states vary in the roles played by these different constituencies, and therefore, in the relevance of their knowledge or opinions. However, a general approach was needed that could be applied to all states, and it is believed that the one chosen is relatively fair and accurate. These procedures were compared with the approach of giving all of the different constituencies the same weight. Results were similar to those produced by the alternative weighting scheme discussed above, which is believed to have greater face validity.

1.2.6 Data Analysis: Cross-tabulations and Tests of Statistical Significance

As previously indicated, response frequencies were tabulated by the constituency of which the respondent was a member. In addition, a second classification of types of respondent was developed for analytic purposes. This second classification scheme combined all of the state level respondents (i.e., the Chief State School Officer, the Assessment Director, the Curriculum Director, the State Board of Education Chair, the Governor's Education Aide, and State Legislative staff) into a single category. It created three types of respondents: State, District, and Association staff. Response frequencies were also cross-tabulated by these types of constituency.

The distributions of response frequencies were reviewed to allow the dichotomization of many of the survey items. For example, four-point importance scales were dichotomized into "Very important/Not very important" or "Moderately or very important/Not moderately or very important" categories, depending on which cut-points produced approximately equal numbers of respondents per category. Chi-squared tests were conducted for each item in both sets of cross-tabulations, to determine if responses were associated with particular constituencies. They were also conducted on the dichotomized items.

If these chi-squared tests indicated an association between type (constituency) of respondent and item responses, further analyses were conducted to identify the types of responses associated with specific constituencies. Comparisons between the proportions of different respondents responding in a specific manner (i.e., "Strongly agree") were performed by calculating z scores in the following manner:

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

where p_1=proportion in group one responding positively to item, q_1= (1 - p_1), n_1=number of respondents in group one, p_2=proportion in group two responding positively to item, q_2= (1 - p_2), and n_2=number of respondents in group two. Similar procedures, employing the square root of ((p*q)/n) to approximate the standard error of a proportion, were used to estimate the probability that the proportion of respondents responding in a specific way to an item was due to chance.

Item 9 (Desire for Information on Skill Areas Within Subjects) was also analyzed by treating the four-point rating scale as if it were an interval scale. Multivariate analyses were performed using the SAS procedure called PROC GLM. This procedure is based on ordinary least-squares (general linear) regression. The model employed in this analysis

treated grade level, subject, type of respondent, and grade level*subject as the independent (control) variables. Subject was nested within grade level; both were nested within type of respondent. These analyses enabled assessment of how important subscale scores were for the different constituencies, the different subjects, and at different grade levels, controlling for other factors.

Different procedures were used to compare responses to sub-items within a question. These procedures were employed to compensate for the fact that the samples of respondents to these items were identical (and not independent). For example, in item 1, respondents were asked to evaluate the importance of four different types of background information. In order to compare the proportions of respondents who deemed each of these factors "very important", the following procedures were employed:

(1) Responses were dichotomized into "Very important (=1)" or "Not very important (=0)" categories.

(2) Within subject difference (comparison) scores were calculated. For example, assume the first respondent said "Instructional practices were not very important". A value of 0 would represent his response to this item. Assume his response to the next item was "School characteristics were very important". A value of 1 would represent this response. The comparison score for these two items is then determined by subtracting the coded response to the second item from the coded response to the first item (i.e., 0 - 1 = -1).

(3) To determine whether respondents felt that instructional practices were more important to study than school characteristics, T-tests for dependent samples were used to test the hypothesis that the mean value of the differences was significantly different from zero.

In order to compensate for errors of inference based upon multiple comparisons, the Bonferroni procedure was employed to correct all significance tests. This procedure corrects the significance level by dividing the normal criterion for statistical significance ($p < .05$) by the total number of contrasts made.

1.3 Focus Groups: Protocol Development

Protocols for focus groups were developed to parallel the structure of the basic survey. That is, all of the items contained in the survey instrument (version III) were translated into topic areas to be discussed in the focus group sessions. This ensured a common content in both the surveys and focus groups (see Appendix C).

The focus group attendees were provided with general information about NAEP. Draft versions of the focus group guides were developed and internally reviewed by AIR and ESSI staff prior to their implementation.

1.3.1 Focus Groups: Participant Recruitment and Operations

During the initial phases of survey development, NCES and NAGB expressed the desire for input on the NAEP redesign from seven other groups in addition to those who would be receiving the constituents' survey. The groups (constituencies) consisted of:

(1) public high school principals,
(2) private high school principals,
(3) elementary school principals,
(4) members of the general and education press,
(5) the general public,
(6) national business organizations which conduct efforts to support and improve schools, and
(7) teachers.

Since it was not feasible to survey these groups, focus groups were held from March 9, 1997 through May 7, 1997 to discuss as many of the relevant issues covered in the questionnaire as possible.

Because only one focus group was conducted for each of these constituencies, it is important to point out that the groups' opinions are not necessarily representative of their larger populations. Thus, the comments from the sessions should not be generalized to all members of each constituency. However, such feedback does provide valuable input from the various groups of participants. The details of each focus group session are described below and in the following sections.

With the exception of the general public and the business organization focus groups, focus groups were conducted during annual conventions to involve participants from different geographic regions. During all sessions, a light meal was provided for participants. Sessions were held in a small conference room (with the exception of the business leaders group) and comments were tape recorded for later report preparation. Notes were taken by ESSI staff at all sessions. The audio tapes were compared with the recorded notes to improve the accuracy of the report. Further details on the methods used for note preparation and analysis are detailed in section 1.3.3.

To solicit input from the research community, the NAEP Constituents' Survey was sent to all current American Education Research Association (AERA) and National

Council on Measurement in Education (NCME) officers and their immediate predecessors. Since their concerns were expected to be very different from those of the constituencies involved with the implementation of NAEP, these results were analyzed separately. They are reported in the Focus Group section.

Principals' Focus Groups. Focus groups were held in conjunction with the annual conventions of the National Association of Secondary School Principals (NASSP) and the National Association of Elementary School Principals (NAESP) in order to meet with principals from across the country. Executive directors at both associations were contacted in advance, and permission was obtained to hold the focus groups during the conferences at lunch time.

Mailing lists of conference attendees were provided by each organization. NASSP provided a list of 80 public school and 64 private school principals who would be attending the conference. From the lists, 37 principals were contacted by phone and invited to participate in the focus groups. Nine public school and nine private school principals agreed to participate in the two focus group sessions (one for private school and one for public school principals). Of the 19 that declined to participate, five said they were unsure about participating and needed more information about the purpose of the focus group before agreeing to participate. The other 14 indicated they had other commitments or would not be attending the conference on the scheduled day and time.

NAESP provided a random sample of 299 conference attendees, with no designation as to whether attendees were elementary school principals. From this list, 19 participants were contacted by phone. Twelve principals agreed to participate in the focus group. Of the seven that did not agree to participate, one was not a principal, two had other commitments, and four did not call back with a decision in time to participate. Two days before the focus group, one of the twelve principals who originally agreed to participate declined.

After the principals agreed to participate in the focus groups, a letter was sent to them which indicated the purpose, location, and time for the session. In addition, a brochure was enclosed to provide background information on NAEP. This information was sent to participants on the same day they were contacted. Two days before each session, calls were made to remind principals of the meeting place and time, and a phone number was given in case they were unable to attend.

The two focus groups with NASSP participants were both held on Sunday, March 9, 1997 at the Omni Rosen Hotel in Orlando, Florida from 12:30 to 2:00 pm. The focus group at the NAESP conference was held on Monday, April 13, 1997 at the Marriott Rivercenter Hotel in San Antonio, Texas from 12:30 to 2:00 pm.

Media Focus Group. Members of the general and education press were invited to attend a focus group during the Education Writers Association (EWA) Meeting which was held in Washington, D.C. The coordinator of the EWA meeting was initially contacted, and she provided an address list for 144 members who would be in attendance. Letters of invitation were sent to all members from the mailing list. Eight EWA members contacted ESSI to participate in the focus group. In addition, calls were made by ESSI to writers at a large national publication and a large education publication. One writer from each of these periodicals agreed to attend the focus group session, for a total of ten participants. The media focus group was held at the Radisson Barcelo Hotel in Washington, DC on May 1, 1997 from 12:00 to 2:00 p.m.

General Public Focus Group. Because the general public differs considerably from the other constituent groups both in its knowledge of assessments in general and of NAEP in particular, it was decided that an experienced general public focus group moderator should lead this focus group. Westat, Inc. of Rockville, MD, an ESSI organization, was contacted, and a special focus group protocol was developed to make the discussion topics relevant and interesting to the public. Westat also developed a focus group screener and contacted members of the public in the Washington, D.C. metropolitan area from their extensive database to identify twelve potential participants. Due to cost constraints, it was not feasible to conduct focus groups with the public in other geographic areas.

The focus group was held at Westat's focus group facility from 6:00 pm to 8:00 pm on May 6, 1997. The facility consists of a small conference room and an observation room with a one-way glass partition. As participants arrived at Westat, ten individuals were selected by the moderator to participate in the session. All individuals who arrived at Westat to participate were paid $50.

National Business Organizations Focus Group. For the focus group with members of national business organizations that conduct efforts to support and improve schools, three leaders of business-in-education programs were invited to attend a luncheon focus group to discuss the redesign of NAEP. All three agreed to participate. The individuals were selected because it was felt that their organizations represented the general business community's interests in education. The participants met with ESSI staff at the Oval Room Restaurant in Washington, DC from 12:00 to 1:30 p.m. on May 7, 1997. A table was reserved in a corner of the restaurant, and lunch was served during the meeting.

Teacher Leaders Focus Group. In order to gather the opinions of teachers on the redesign of NAEP, ESSI contacted the American Federation of Teachers (AFT) to determine whether they would be conducting any meetings in the Washington, D.C. area.

The meeting of their K-12 Program and Policy Council (K-12 PPC) was taking place in early May. (The Program and Policy Council is a national leadership group of AFT members.) ESSI prepared materials which were faxed by AFT to all 38 members inviting them to attend a NAEP focus group during the meeting. Five members called ESSI to say that they would participate in the focus group. The session was held on May 7, 1997 from 4:30 to 6:00 pm at the Mayflower Hotel after the regular daily AFT meeting sessions were completed.

Education Researchers' Survey. AERA was contacted and provided a list of 40 past and current AERA and NCME officers. Version III of the NAEP Constituents' Survey (which was previously sent to education associations and superintendents) was mailed to all 40 officers on April 21, 1997. A return date of May 1, 1997 was included on the survey. Nonrespondent follow-up took place throughout the month of May. Ultimately, surveys were received from 32 of the 40 AERA and NCME officers -- an 80 percent response rate.

1.3.2 Demographics of Focus Group Participants

In total, 46 individuals participated in the focus groups. The individuals who attended the sessions came from 19 different states and the District of Columbia. This section describes the participants in each focus group.

Principal Demographics. Nine principals attended the NASSP focus group sessions. There were four participants in the private school group, one of whom was a public school principal who was accidentally included in NASSP's private school list. Five public school principals attended the public school group. Tables 1-3 and 1-4 indicate the groups' compositions.

Table 1-3. NASSP Public School Group

Race/ Ethnicity	Gender	School Type	Enrollment	Grades	State
white	male	public	1,050	6-8	Massachusetts
white	female	public	less than 400	9-12	New Jersey
white	male	public	848	9-12	New Jersey
white	male	public	1,582	didn't mention	Massachusetts
white	female	public	didn't mention	high school	New Jersey

Table 1-4. NASSP Private School Group

Race/ Ethnicity	Gender	School Type	Enrollment	Grades	State
white	female	private	750	9-12	Illinois
white	male	private	636	preK-12	Louisiana
white	male	private	1,550	preK-12	Tennessee
white	female	public	2,200	9-12	Florida

The NASSP focus groups were comprised of five males and four females. One of the private school participants was a principal of a Catholic school, and two principals worked at "unit" schools which serve students from prekindergarten to twelfth grade. The school enrollment size varied from less than 400 students to over 2,000 students. Although a random sample of principals was drawn by NASSP and was systematically sampled by ESSI, the public school group consisted solely of participants from Massachusetts and New Jersey.

Information on student demographics was collected from four of the public school principals. Two of the schools were located in relatively affluent areas, and two were located in middle class, "blue collar" communities. Two of the communities were in rapidly growing areas.

At the NAESP conference, eight public elementary school principals and one director of principals attended the focus group. Table 1-5 provides the demographics of the focus group.

Table 1-5. NAESP Group

Race/ Ethnicity	Gender	School Type	Enrollment	Grades	State
white	female	public	725	didn't mention	Georgia
white	female	public	600	didn't mention	Texas
white	female	public	864	preK - 6	Texas
white	female	public	didn't mention	preK - 4	Texas
white	male	public	26 schools		Texas
white	female	public	didn't mention	preK - 2	Georgia
white	female	public	450	preK - 5	Connecticut
white	female	public	didn't mention	preK - 5	Texas
white	female	public	326	K - 3	Pennsylvania

Eight females and one male participated in the focus group. The schools varied in the grade levels served and enrollment. Although a random sample of participants was sent from NAESP, the focus group consisted of five members from Texas. This may have resulted from the calling schedule, which took place over the three weeks most likely to be the spring break time period of other school systems.

The schools varied in socioeconomic status, with six schools in the middle to upper socioeconomic level, and three in the lower socioeconomic level. Minority student racial composition within schools ranged from 10 percent (Korean students) to 48 percent (38 percent Hispanic and 10 percent Black). One school was located in a rural community; the rest were located in suburban areas. Two principals mentioned that they worked in Title 1 schools.

Education Writers Participants. A total of ten education writers attended the focus group which was held concurrent with the Education Writers Association Meeting. Table 1-6 details the background of the group members.

Table 1-6. Media Group

Race/ Ethnicity	Gender	Publication Type	Circulation
white	male	education	national
white	male	general	local
white	female	general	large city
white	female	education freelance	---
black	female	education	daily national
black	female	general	large national
black	female	university publication	---
white	female	general	small county
white	male	education	education testing agency
white	female	general	large city

There were five education writers from the general press and five writers from the education press. The group participants came from seven different states and served a variety of audiences. Three of the publications had large circulations.

General Public Participants. Ten members of the general public took part in the focus group at the Telephone Research Center at Westat, Inc. Table 1-7 provides additional details about the participants.

Table 1-7. General Public Focus Group

Age	Gender	Childrens' Ages (< 21)	Race/ Ethnicity	Education	Occupation
43	female	19 and 17	black	technical school	personnel manager
57	male	17	black	2 yr. of college	computer specialist
52	female	None	white	BA	real estate agent
37	female	13 months	Hispanic	MA	social worker
49	female	None	white	HS diploma	customer service
45	female	None	white	2 yr. of college	office manager
54	female	None	white	BA	instructional assistant
51	male	None	white	BA	Director, HR
51	male	12 and 10	white	2 yr. of college	paralegal
38	male	16, 9, 18 mo.	black	BA	communications

The six women and four men ranged in age from 37 to 57. Three participants were African-American and one was Hispanic; the rest were white. One participant had a high school education only, four had some college or technical education, four had bachelor's degrees, and one person had a master's degree. Only four participants had children of elementary or secondary school age, but all but one participant was a parent or stepparent.

National Business Organization Participants. All three participants in the group were females who worked as directors of educational programs in large business organizations in the Washington, DC area.

Teacher Participants. A total of five AFT K-12 PPC Committee members attended the focus group. One of the members did not arrive until 5:45, when the meeting was almost finished. Table 1-8 indicates the group's composition.

Table 1-8. AFT K-12 PPC Committee

Race/ Ethnicity	Gender	Region	Teaching Experience
white	male	Midwest	15 years, secondary level (history)
white	female	East	22 years, middle and secondary level (language arts)
white	female	Midwest	25 years, all levels (5 areas)
black	female	South	16 years
white	female	Northeast	arrived late, no information

The AFT group was composed of four females and one male. Two of the participants who arrived did not initially call ESSI to indicate that they would be attending. Also, two members who said they would participate did not attend the session. The focus group participants represented states that varied greatly in their rank on previous NAEP assessments. Two of the members expressed that they did a considerable amount of work in assessment and accountability, and one was on a board of trustees for a university.

1.3.3 Analysis of Focus Group Responses

The focus group protocol was developed to cover the same topics that were included in the NAEP Constituents' Survey. After each focus group, summaries of responses were prepared. These summaries were based on notes taken during the session and a review of the audiotape of the session. Responses were grouped by content and tone, and in one case, categorical responses were counted for each session. In this way, the spectrum of feedback from the group could be accurately summarized. Any conclusions which were made about a focus group were based on a consensus of responses by members.

2.0 SURVEY RESULTS

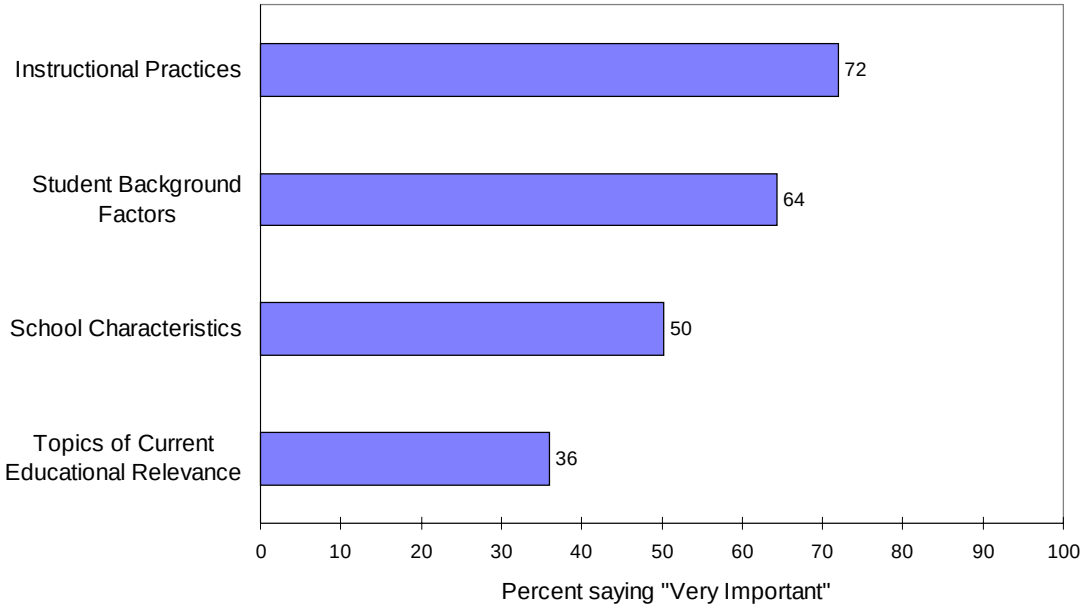
Survey results are presented for each questionnaire item. All programmatically meaningful, statistically significant relationships are reported. Statistically significant relationships where a higher rate of "Don't know" responses is associated with a specific constituency (e.g., State Legislative Committee staff) are not reported. This is because such findings are felt to be of minimal programmatic interest. When responses were homogeneous across the different constituent groups, with few or no significant difference between groups, figures present means for the entire sample of respondents.

2.1 Background Information -- Emphasis

NAEP collects school, teacher, and student background information to permit investigation of relationships between these factors and student performance on its assessments. Since it is not feasible to collect extensive amounts of background data, respondents were asked how important they felt each of four different areas were for NAEP to study in order to prioritize the types of background information that is collected. Members, on average, had comparable opinions about the importance of these areas. Accordingly, responses are summarized as the overall percentages of respondents who felt each area was "very important" for NAEP to study.

Both *Instructional Practices* and *Student Background Factors* were felt to be significantly more important for NAEP to study than either *School Characteristics* or *Topics of Current Educational Relevance* ($p < .01$). Nearly three-quarters (72 percent) of the respondents felt that *Instructional Practices* were very important and nearly two-thirds (64 percent) felt that *Student Background Factors* were very important for NAEP to study (see figure 2-1).

Figure 2-1. Background Information - Emphasis: How important are each of the following for NAEP to study?



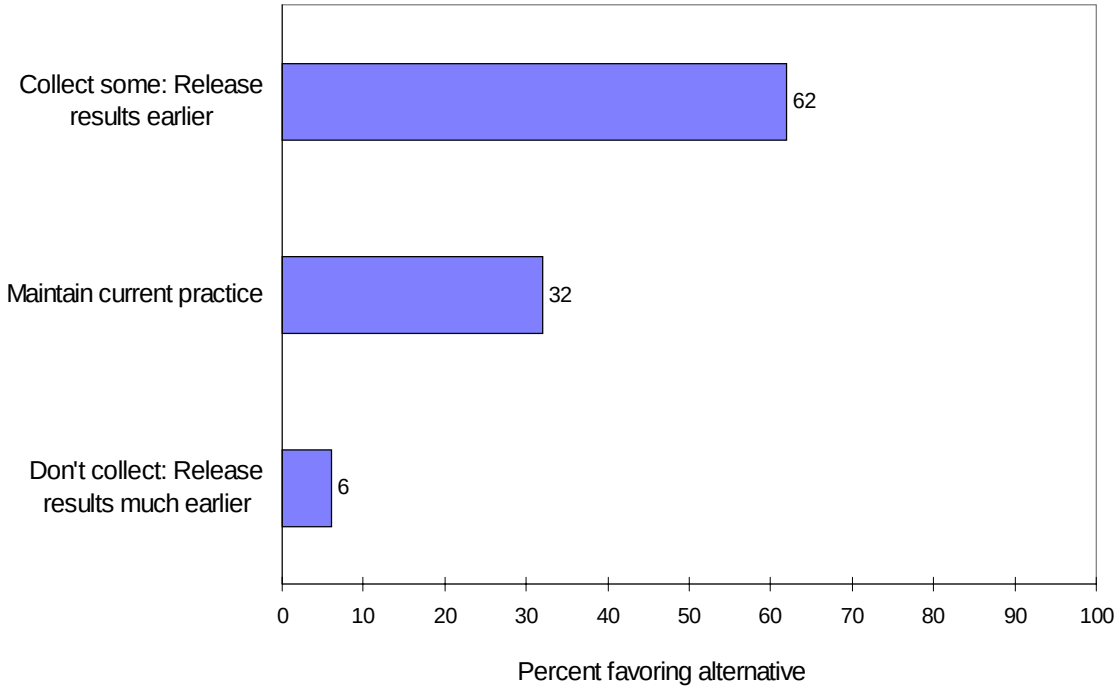
Only about one-third (36 percent) of the respondents felt that *Topics of Current Educational Relevance* were very important for NAEP to study. Such topics might include portfolio assessment or charter schools, for example. Some individuals indicated they were unsure about the meaning of this factor. One respondent commented: "Unsure of how you would use these topics. Our responses would depend on your plans." These feelings contributed to the area's comparatively low evaluation. However, in addition to the 36 percent of respondents who felt this was a "very important" area for NAEP to study, another one-third of the respondents felt this was a "moderately important" area for NAEP to study. In other words, over two-thirds (68 percent) of the respondents felt that the least important suggested area was at least moderately important for NAEP to study. Very few respondents felt that these areas were not worthy of investigation.

2.2 Background Questions - Impacts on Release of Results

NAEP has to analyze responses to all of the questions about student, teacher, and school background to enable its analyses of student achievement to be conducted. These analyses take time. Respondents were asked to consider the value of these items in light of their impact on when results could be released. About two-thirds (62 percent) were willing to cut back on the amount or frequency of background information data collection to allow an earlier release of NAEP results ($p < .001$). Almost no respondents (just 6

percent) felt that a much speedier release of results would justify elimination of these items ($p < .001$, see figure 2-2).

Figure 2-2. Background Information: Impacts on Release of Results. (Trade-offs between collecting background information and time of release of results)



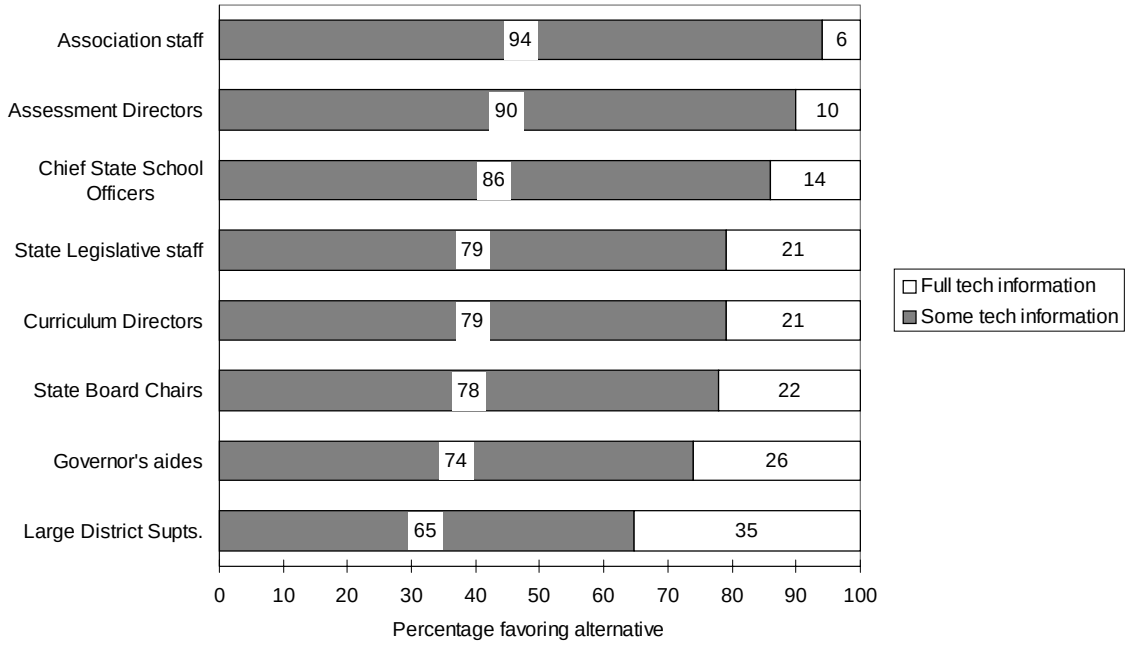
This question elicited more comments than any other (see Appendix B). About ten respondents suggested another alternative: an early release of results with limited background information, with full background information to follow. This compromise would allow for greater timeliness without sacrificing technical detail.

2.3 Technical Documentation of NAEP vs. Timeliness of Reports

NAEP's high standards include the provision of extensive documentation about the technical characteristics of the assessment. Preparing this extensive documentation is a time-consuming process, impacting when results can be released. Respondents were overwhelmingly in favor (78 percent vs. 22 percent, $p < .001$) of cutting back on the amount of technical documentation that is provided so results could be released about six months sooner. Several respondents also suggested an earlier release with limited documentation, followed by a subsequent release of the full technical documentation.

Although respondents were strongly in favor of an earlier release with less technical documentation, these feelings were stronger in certain constituencies. In comparison with Superintendents of large school districts, both Association staff and State Assessment Directors were stronger supporters of cutting back on the amount of technical documentation to permit an earlier release of results ($p < .01$, see figure 2-3).

Figure 2-3. Technical Documentation vs. Timeliness: Preference for earlier release vs. full technical documentation



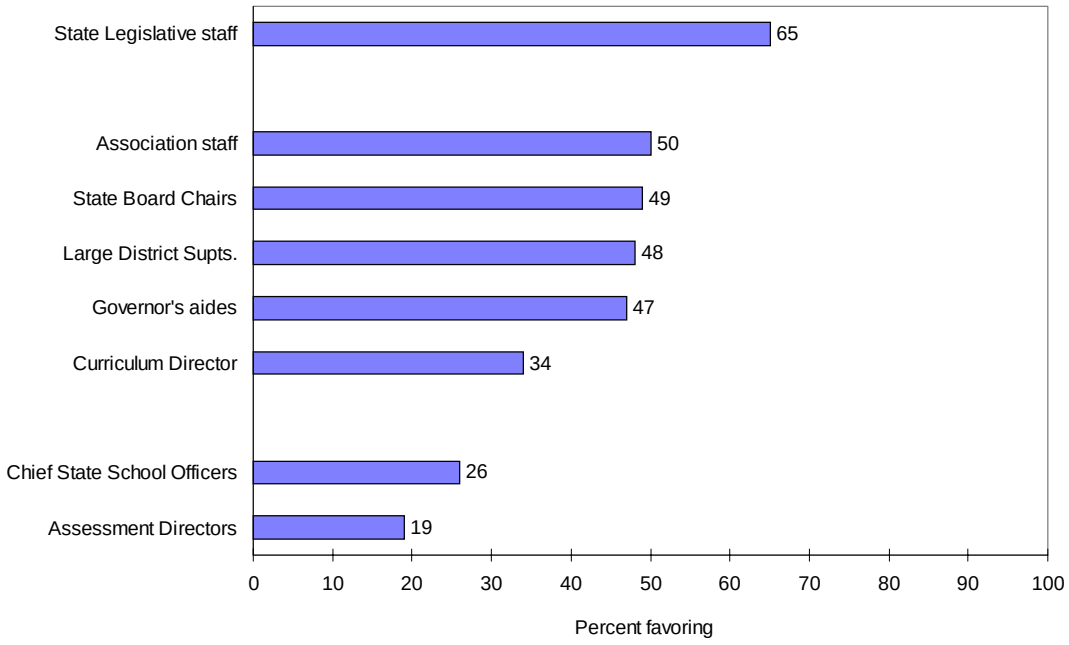
2.4 Including a Parent Survey

Some of the background information collected by NAEP can be provided most accurately by parents. However, the collection of data in this way would represent an added burden on schools, as well as raise political issues. Considering the trade-offs (the value of the information versus its costs), respondents were opposed to the idea of a parent survey for NAEP ($p < .001$). Only 41 percent of the respondents felt NAEP should either "probably have a parent survey" or "definitely have a parent survey."

Representatives of only one constituency -- State Legislative staff (that is, education advisors to state legislature committees) -- were in favor of a parent survey. Significantly fewer Assessment Directors ($p < .01$) or Chief State School Officers

($p < .05$) than Legislative staff favor this data collection approach. In general, the constituencies that would be most strongly involved in the implementation of a parent's survey (and most subject to criticisms of such an approach) were the constituencies that were most strongly opposed (see figure 2-4).

Figure 2-4. Parent Survey: Opinion about NAEP including a questionnaire filled out by parents



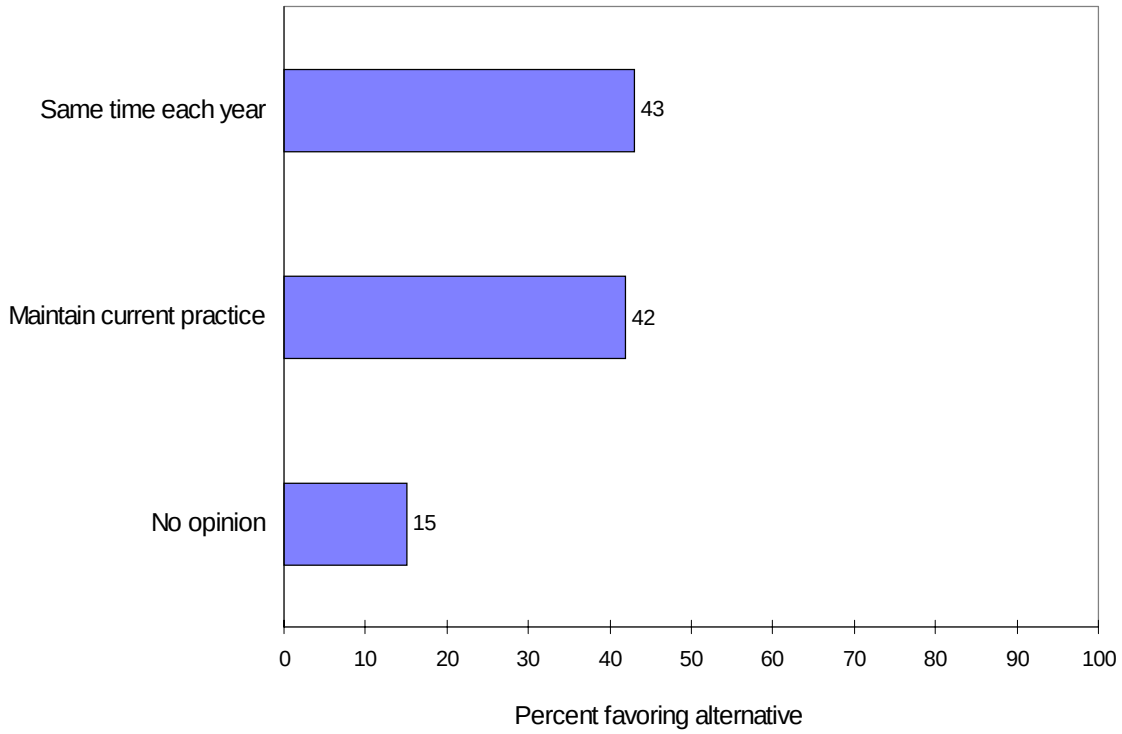
Some respondents pointed out logistical problems that would be associated with a parents' survey. One claimed that 85 percent of the parents in his district were not English speakers. Another respondent feared that it would discourage school participation in NAEP. Conversely, many argued for the potential value of such a survey. Compromises of occasional, infrequent parent surveys and brief, focused parental surveys were suggested.

2.5 Reporting: Schedule for Release of Results

In the past, NAEP reports have been released when they were available. There was no fixed schedule. Respondents were asked whether they would prefer that a schedule for release of results be established or whether the current approach of releasing results when available was acceptable. Since there were no significant differences between constituencies with respect to their preferences, results are presented as overall percentages.

About one-sixth (15 percent) of the respondents indicated that they had no opinion or that it didn't matter to them. The remaining respondents were nearly equally divided among those favoring a set schedule versus those favoring the current approach (see figure 2-5).

Figure 2-5. Schedule for Release of Results

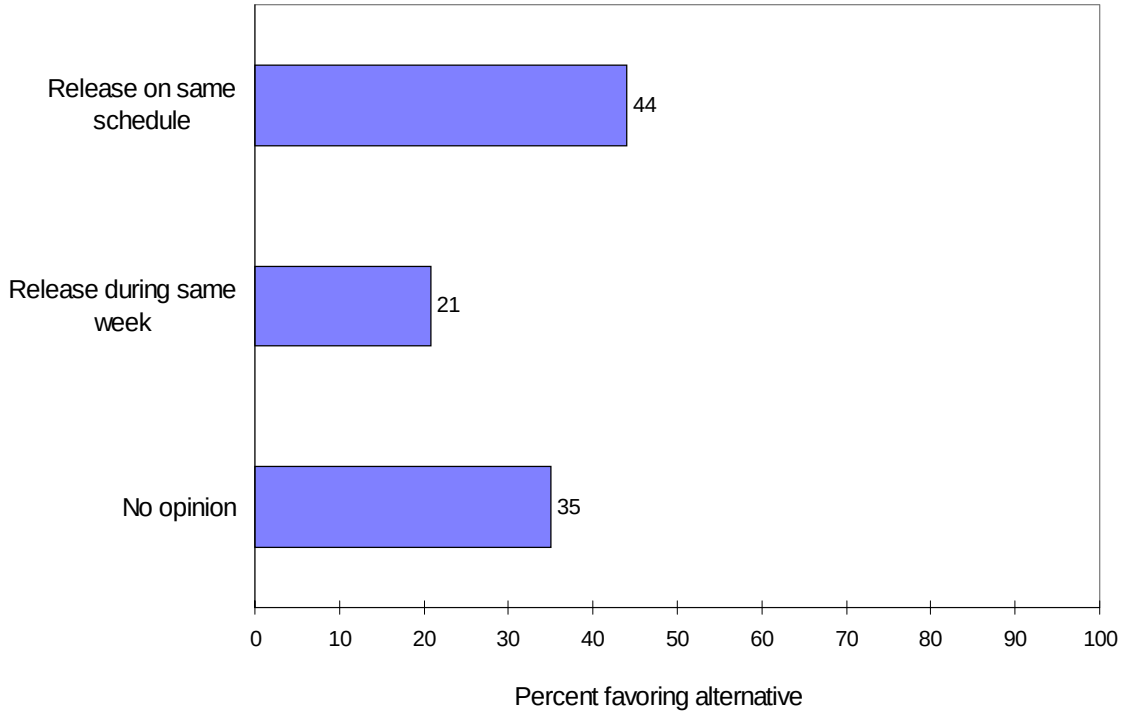


Those favoring a set schedule presented powerful arguments for the practice. It would allow them to prepare their membership (or agency) for the release. Some also felt that a set schedule would increase NAEP's impact. Conversely, other respondents commented that a set schedule could delay release of results or result in lower quality reports.

If changes were to be made in the schedule, a variety of different schedules are worthy of consideration. NAEP might be released on the same schedule each year (e.g., always around the start of February and May). Or, there might be a specific period each year (such as a week) when all of a given year's NAEP data are released. About one-third (35 percent) of the respondents had no opinion or felt that it didn't matter which of these options were selected. Of those with an opinion, twice as many respondents favored the

release of NAEP results on a specific schedule each year rather than releasing all of the year's results during a specific week (see figure 2-6).

Figure 2-6. Schedule for Release of Results: Release on consistent schedule or all at once



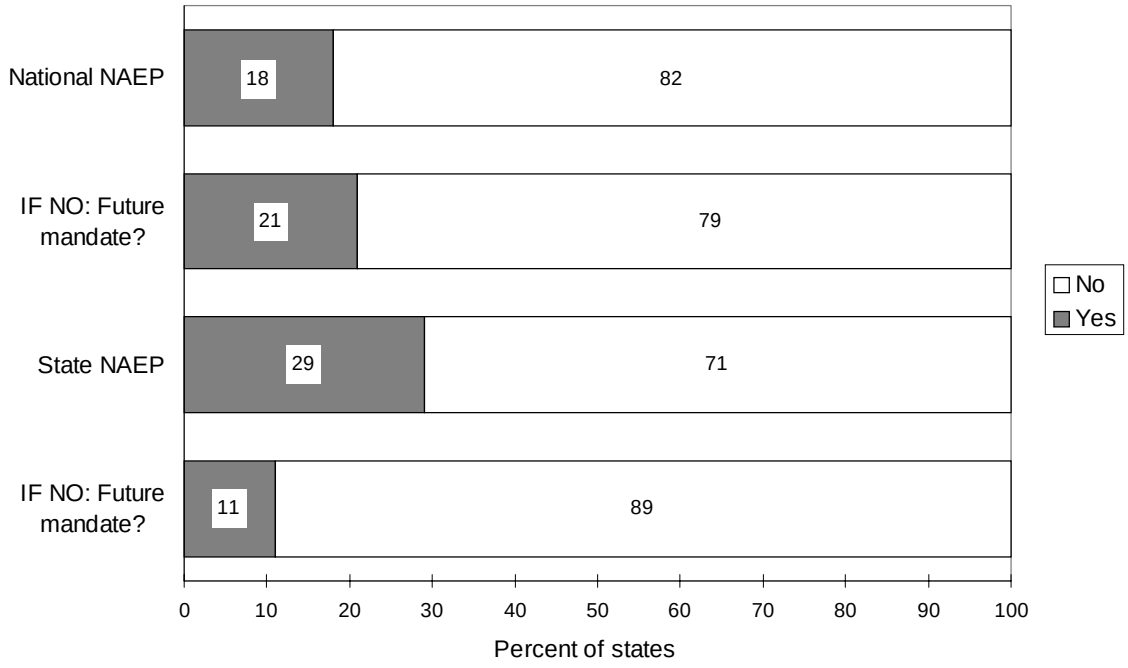
2.6 State Mandates for Participation in NAEP

During the most recent assessments, students in thousands of different schools participated in the national and state NAEPs. In some cases, these schools were required to participate by their State Education Agency. In other cases, participation was voluntary.

Most states do not require (mandate) participation in either the state (71 percent, $p < .001$) or national NAEP (82 percent, $p < .001$). When states mandate participation, they are more likely to do so for the state NAEP than the national NAEP ($p < .05$). These policies are unlikely to change in the near future. About 79 percent of the states that do not mandate participation in the national NAEP and 89 percent of the states that do not mandate participation in the state NAEP do not anticipate mandating participation in the

future (see figure 2-7)[5].

Figure 2-7. State Mandates for NAEP Participation: Current and Future Plans



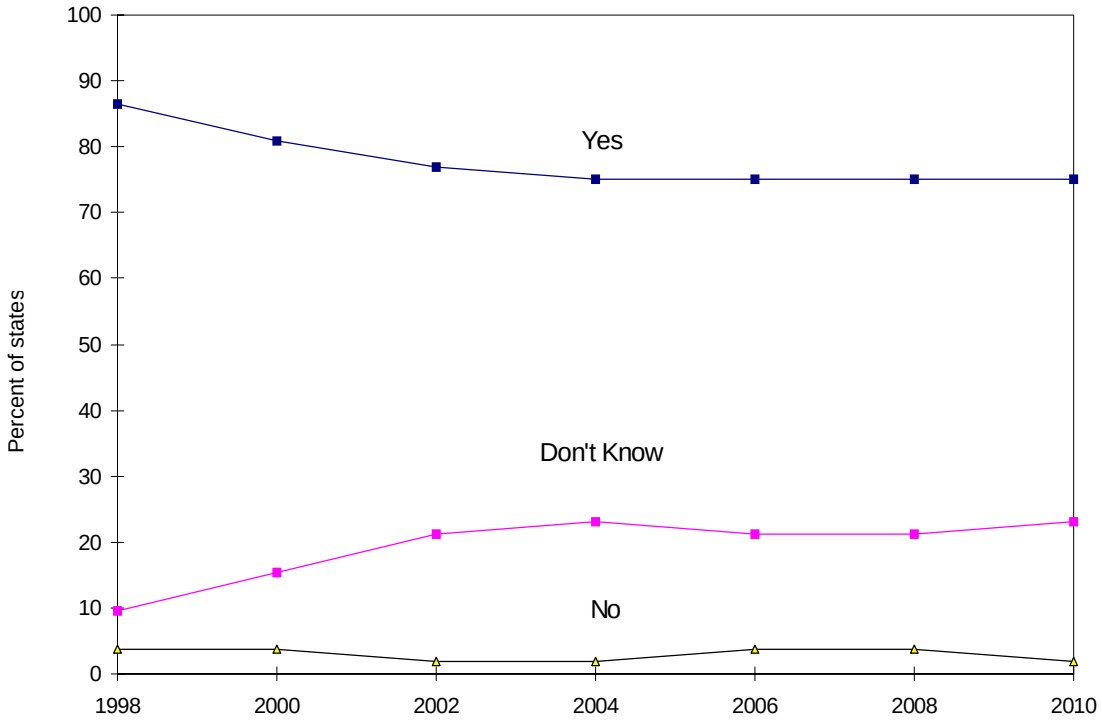
A few respondents indicated there was no need for mandating participation since participation rates in their states were high already. Others responded to our request for suggestions for enhancing district and school participation. These respondents often noted that schools and districts were not being compensated for participation. Some suggested they be provided with either district- or school-level scores. Several others indicated that they would not be at all averse to compensation for the costs associated with NAEP administration.

[5] The question, "Does your state mandate participation in the state-level (or national-level) NAEP?" is a question of fact rather than an attitudinal item. It was felt that the State Assessment Director would be the most knowledgeable person to answer this factual question. Accordingly, responses to this item are those of the State Assessment Director only. (There was nearly perfect agreement between the State Assessment Director's response and the modal state respondent's response to this item. The only inconsistencies occurred when the modal response for the state was "Don't Know" and the Assessment Director's response was affirmative.)

2.7 Subjects Assessed at the State Level

The proposed state NAEP's testing content and schedule for the years 1998 to 2010 was provided to state respondents. They were asked to indicate whether or not they thought their state would participate in each of the listed years. Only one or two states indicated they did not anticipate their state's participation in any of these years. Although uncertainty about participation seemed to increase over time, three-quarters of the states thought they would participate in every state NAEP through the year 2010 (see figure 2-8).

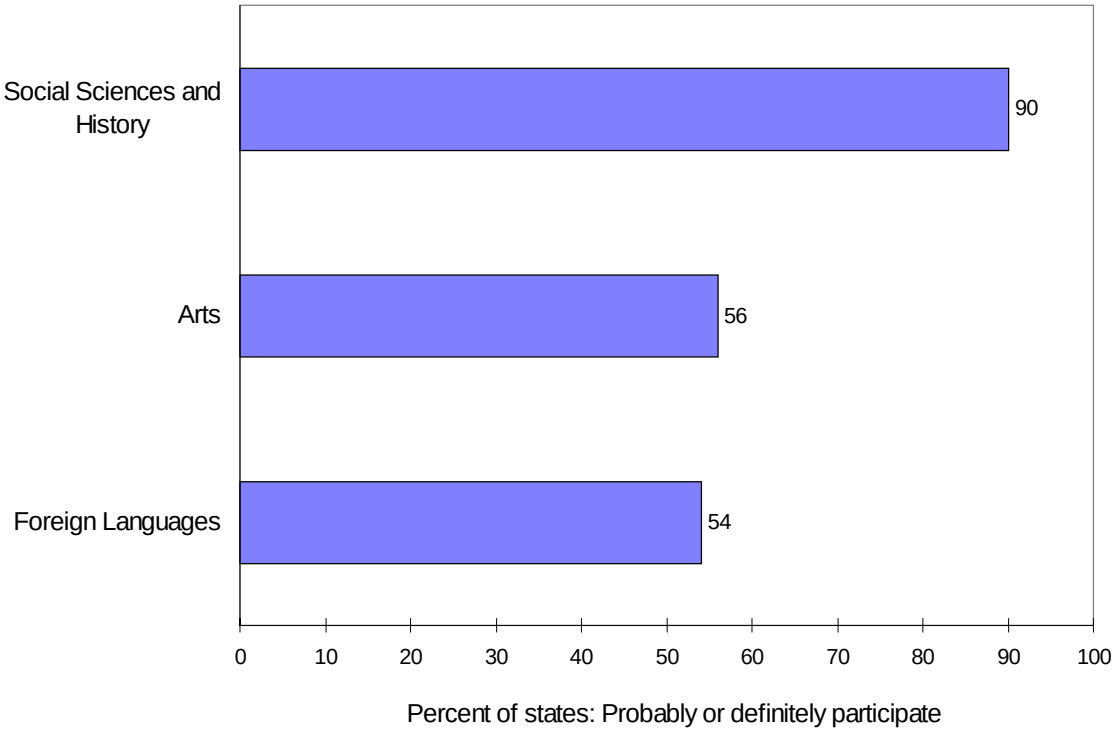
Figure 2-8. Probable Participation in State NAEP: 1998-2010



Some respondents in small states used this item as an opportunity to reiterate their concerns that extensive testing could affect too many, or even every, school in their state, threatening the probability of participation. Others said that costs could threaten participation and that participation could be enhanced if student-level data were made available. On a more positive note, several respondents affirmed their intentions to participate. One proclaimed, "We value our participation!"

Respondents were also asked about their state's interest in participating in assessments in subject areas not currently assessed. These subject areas included social sciences and history, arts, and foreign language. There was substantial interest in social sciences and history assessments: 90 percent of the respondents indicated they were either probably interested (68 percent) or definitely interested (22 percent) in participation.[6] Interest in this area was much greater than interest in either arts assessments or foreign language assessments ($p < .001$, see figure 2-9).

Figure 2-9. Interest in Participating in State NAEP in Different Subjects



Several other areas were suggested for consideration for future state NAEP assessments. The most frequently suggested area was health (or health and safety). Other areas suggested included physical education, technology/computer literacy, biology, algebra, and vocational education.

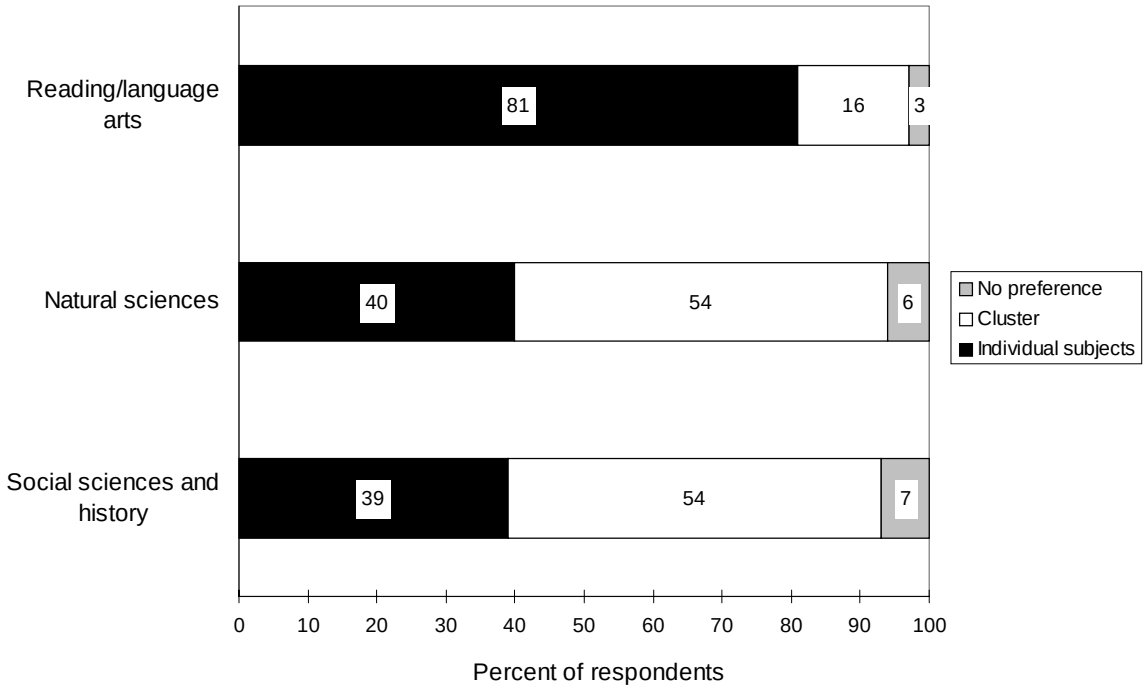
[6] These data summarize the responses of individual respondents rather than responses aggregated at the state level. All other summary statistics for this item represent aggregated state responses. (It is coincidental that 90 percent of the individual respondents and 90 percent of the aggregated state responses indicated that their state probably or definitely would participate in social sciences and history assessments.)

2.8 Assessing Subject Areas in Combinations or Individually

Assessment scores can be measured and reported for general subject areas (such as social sciences or natural sciences) or for separate component areas within these disciplines. There are costs (increased burden) associated with reporting scores for the separate component areas within each discipline. There are also benefits associated with the greater level of detail that can be provided. Respondents were asked about their preferences for measuring and reporting results for social sciences and history, natural sciences, and reading/language arts: whether they would prefer to have scores in each of these areas reported for each individual (component) subject or whether they would prefer the scores be measured and reported as a cluster. With cluster reporting, both an overall "cluster" score (i.e., natural sciences, reading/language arts) and individual (component) subject scores (i.e., earth sciences, life sciences, physical sciences, reading, writing) are reported. However, subscale scores (i.e., Reading for Literary Experience, Reading for Information, Reading to Perform a Task) would not be reported. Associations between student demographic characteristics or school factors and overall cluster scores would be possible. However, with cluster reporting, associations between individual (component) subject scores and student demographics or school factors would not be reported.

For reading/language arts, there was overwhelming support for reporting these scores as individual subjects ($p < .001$). Of the 97 percent of respondents who had a preference, five times as many favored reporting of the individual subject (reading and writing) scores as favored cluster (reading/language arts) score reporting. Conversely, for natural sciences, and social sciences and history, respondents who expressed a preference favored reporting scores as clusters ($p < .05$, see figure 2-10).

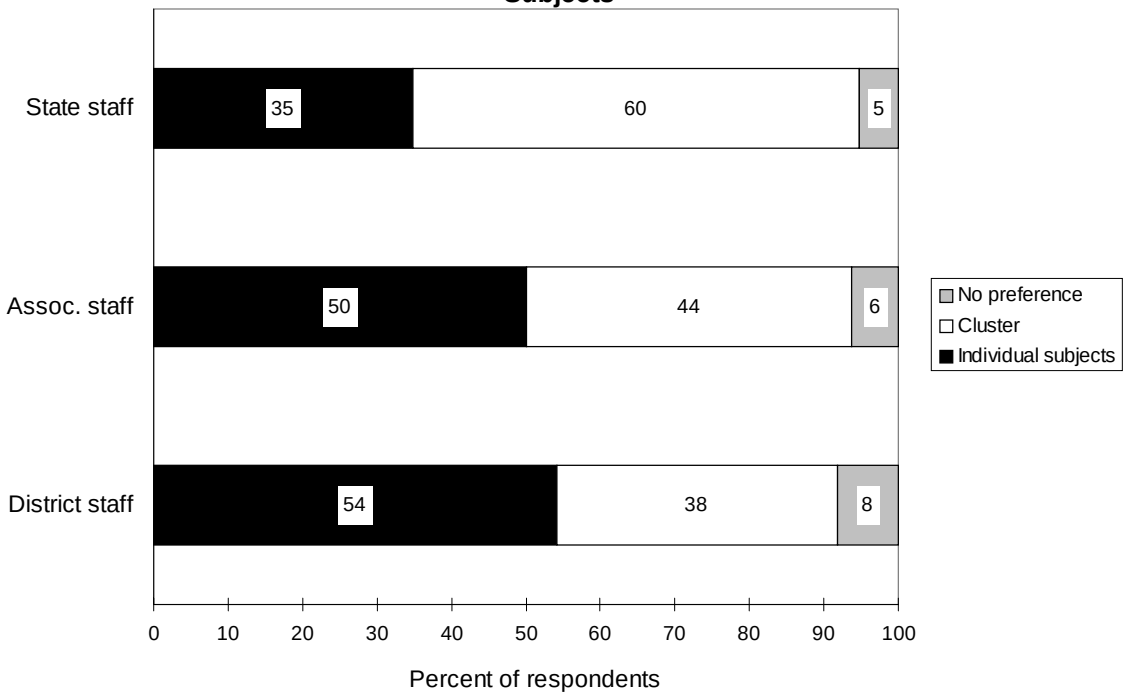
Figure 2-10. Assessing Subject Areas in Combinations or Individually



Arguments in favor of cluster assessment and reporting were related to more integrated or constructivist approaches to student achievement and learning. Those in favor of individual subject reporting noted the benefits of detailed information in addressing problems of poor performance. One respondent suggested a compromise solution: reporting in clusters generally, with individual subject level assessment and reporting occurring every 5 to 10 years.

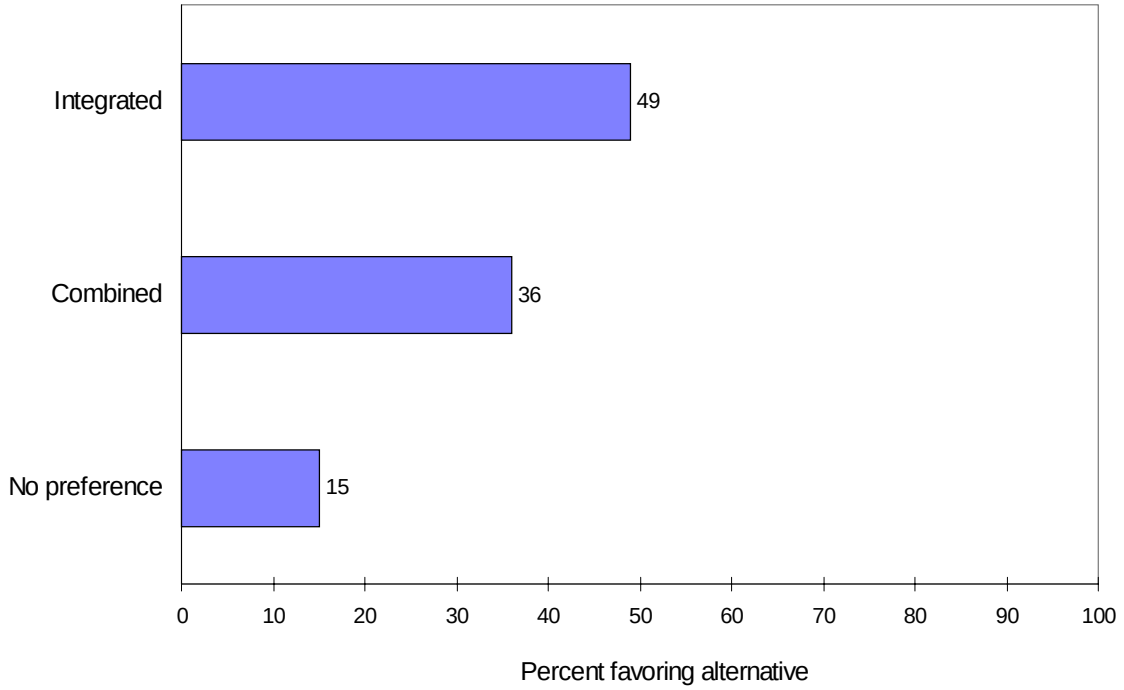
Different constituencies had different feelings about how these subject areas should be assessed. Of the large district superintendents with a preference, 59 percent favored individual subject assessment and reporting in the natural sciences. Only 37 percent of the state-level respondents (State Assessment Directors, State Curriculum Directors, Chief State School Officers, State Board Chairs, Legislative Aides, and Governor's Aides) with a preference favored this kind of reporting for the natural sciences ($p < .01$) (see figure 2-11).

Figure 2-11. Assessing Natural Sciences in a Cluster or as Individual Subjects



Another survey item asked State Assessment Directors and State Curriculum Directors whether they favored assessment and reporting of subject clusters in an integrated (holistic) fashion or their assessment and reporting as free-standing areas. (This item was felt to be too technical to be asked of other constituencies.) About one-sixth (15 percent) of the respondents had no preference. Of the Assessment Directors with a preference, there was a tendency for them to favor an integrated approach (see figure 2-12).

Figure 2-12. Cluster Assessment: Preference for Integrated Assessments vs. Combined Assessments



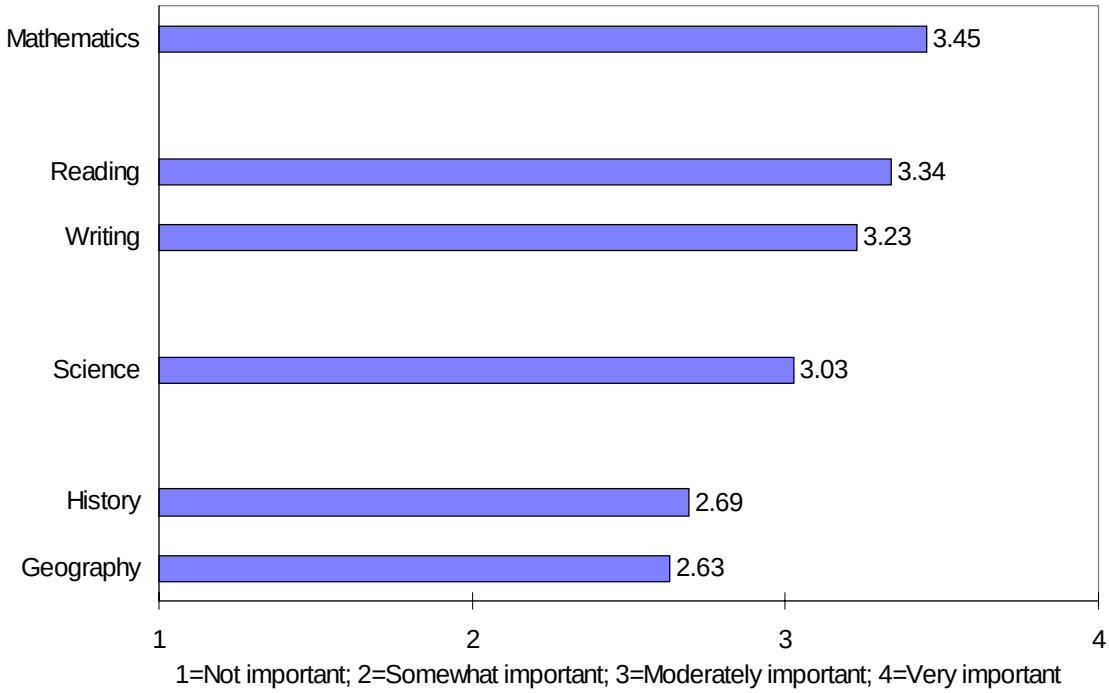
2.9 Desire for Information on Skill Areas within Subjects

NAEP has provided detailed results on sub-areas (skill areas) within the overall disciplines that are assessed. Such detailed information can be of value to educators and researchers. However, since NAEP's resources are limited, the provision of this information means that fewer subjects can be assessed. To permit informed redesign decisions consonant with the needs and desires of the NAEP constituencies, it is useful to identify the skill areas within subjects that are evaluated as being most important to the involved constituencies.

Current NAEP subscales were listed for each of the six different subjects (reading, writing, mathematics, science, geography, and history) and each of the grade levels (fourth, eighth, and twelfth grades) in which NAEP conducts assessments. Respondents were asked to indicate how important it was to have subscale scores for each of these 18 different assessments. To help determine the relative importance of the different subjects and grades, responses were coded on a 1-4 scale (see pages 12-13).

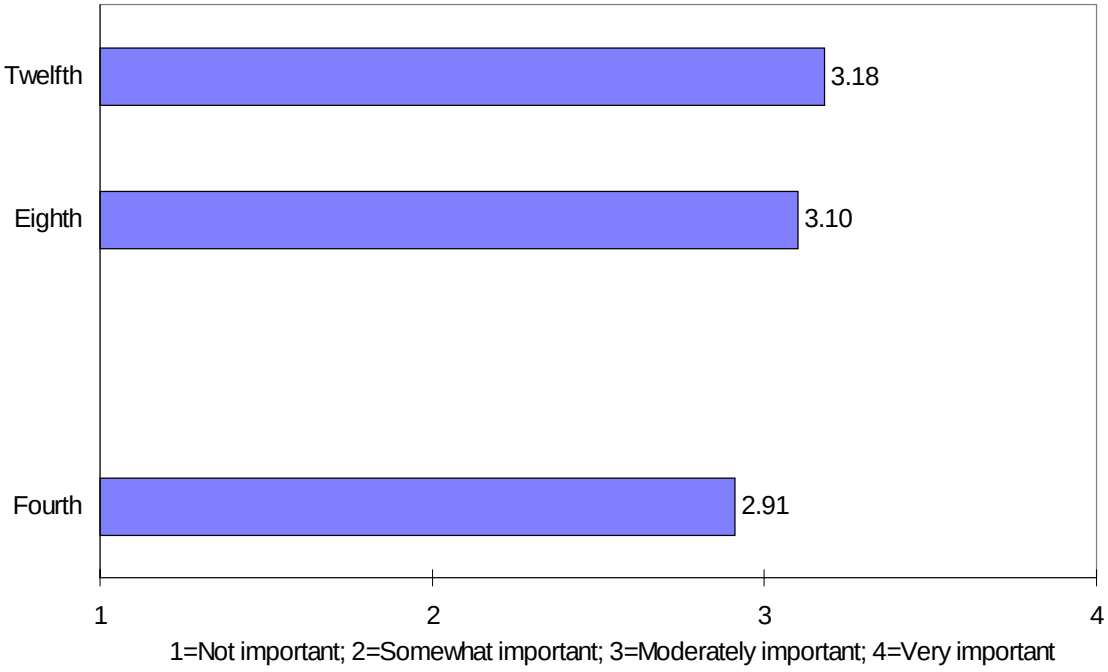
Overall, respondents felt subscale scores were more important in mathematics than in any other area (except reading, $p < .001$.) Similarly, reading subscale scores were more important than in any other area, except for writing and mathematics ($p < .001$). Science subscale scores were more important than subscale scores in history and geography ($p < .001$, see figure 2-13).

Figure 2-13. Importance of Subscale Scores (least-squares means), by Subject Area



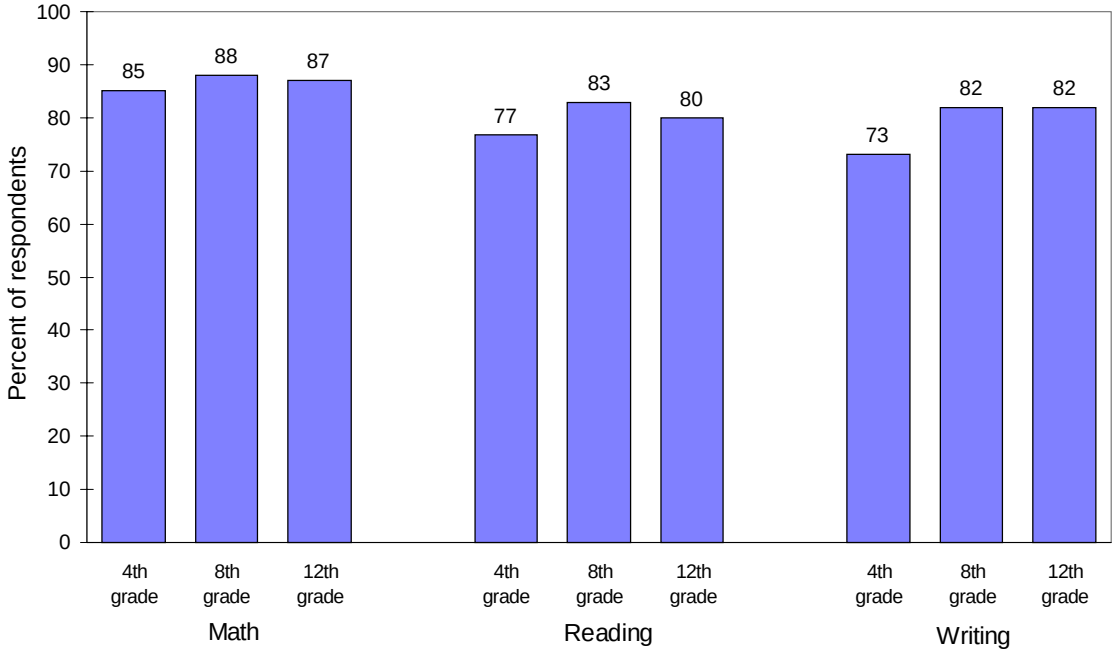
Respondents felt that the importance of subscale scores was greatest for the older (eighth and twelfth grade) students ($p < .001$, see figure 2-14). However, even for fourth graders, respondents felt that subscale scores were moderately important (2.91 on a scale for which 1=Not important, 2=Somewhat important, 3=Moderately important, and 4=Very important).

Figure 2-14. Importance of Subscale Scores (least-squares means), by Grade



Respondents felt that mathematics subscale scores were equally important at all grade levels.[7] They also felt that reading subscale scores were more important for eighth graders than for fourth graders ($p < .01$) and that writing subscale scores were more important for eighth and twelfth graders than for fourth graders ($p < .01$, see figure 2-15).

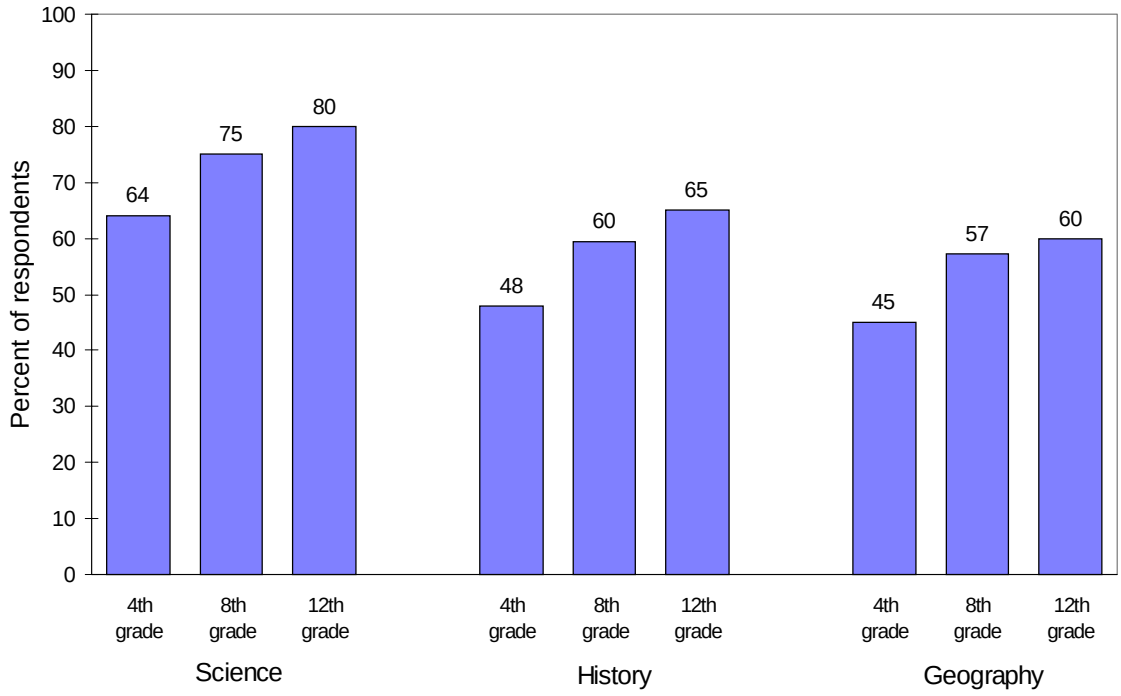
Figure 2-15. Desire for Subscale Information, by Subject and Grade: Percentage of respondents saying information is moderately or very important



[7] Responses to these items (see figures 2-15 and 2-16) are summarized in terms of the proportions of respondents indicating information was "moderately important" or "very important" rather than the proportions responding "very important." This decision was based on a review of the response distributions and employs the criterion discussed on page 12.

Similarly, science, history, and geography subscale scores were more important for eighth and twelfth graders than for fourth graders ($p < .01$, see figure 2-16). Several respondents felt that fourth graders were not advanced enough (in some subjects) to make subscale data useful.

Figure 2-16. Desire for Subscale Information, by Subject and Grade: Percentage of respondents saying information is moderately or very important

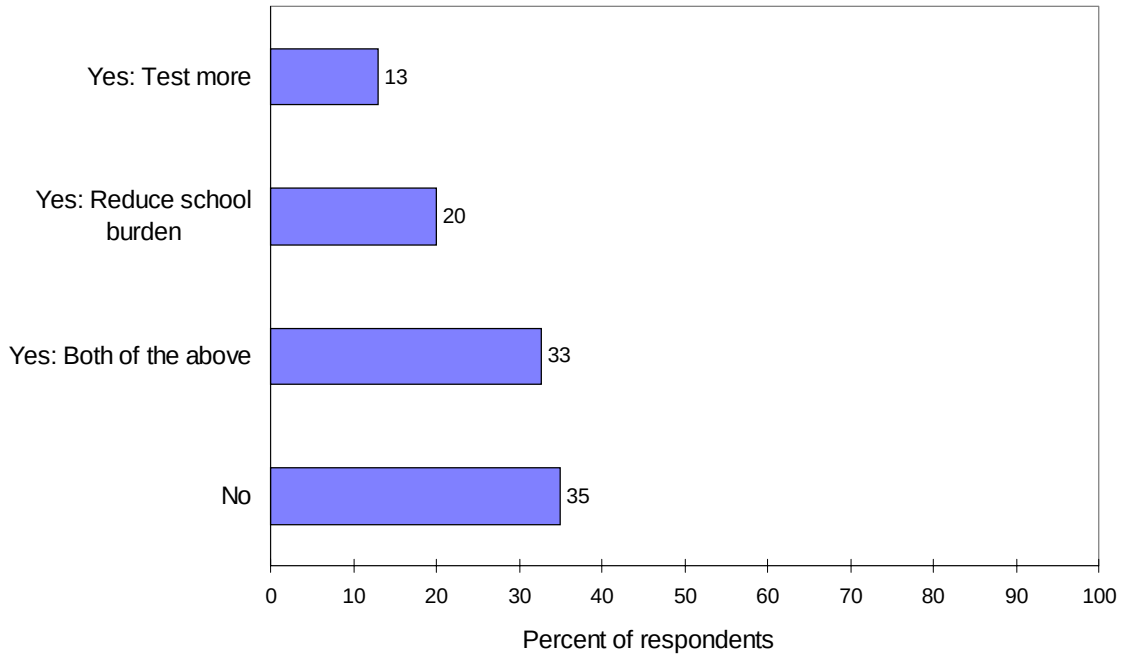


2.10 Frequency of Data Collections

Congress recently authorized NAEP to collect data every year. Previously, NAEP had tested students on a biennial schedule. If a yearly schedule is implemented, it can be implemented in a number of different ways. For example, more subjects could be assessed or the same subjects could be assessed more frequently. Or, the same amount of overall testing could be conducted on a slightly different schedule. For example, instead of assessing mathematics and science in the same year, mathematics could be assessed in one year and science in the subsequent year. This second alternative could reduce the burden on individual schools, since their students would only participate in a single assessment rather than two assessments in a given year.

Respondents favored yearly testing by nearly a two-to-one margin over maintaining the present biennial schedule ($p < .001$, see figure 2-17). However, respondents who were opposed to yearly testing were more inclined to elaborate on their preferences. About two-thirds of the comments provided by respondents demonstrated opposition to yearly testing. Some respondents warned against testing too frequently, since trends take time to develop and "small blips are of little interest". Others expressed the general concern that testing takes time away from teaching activities. And, many were concerned with the burden placed on schools.

Figure 2-17. Annual Data Collections: Should NAEP collect data each year, and why?



2.11 Linking NAEP to International Assessments

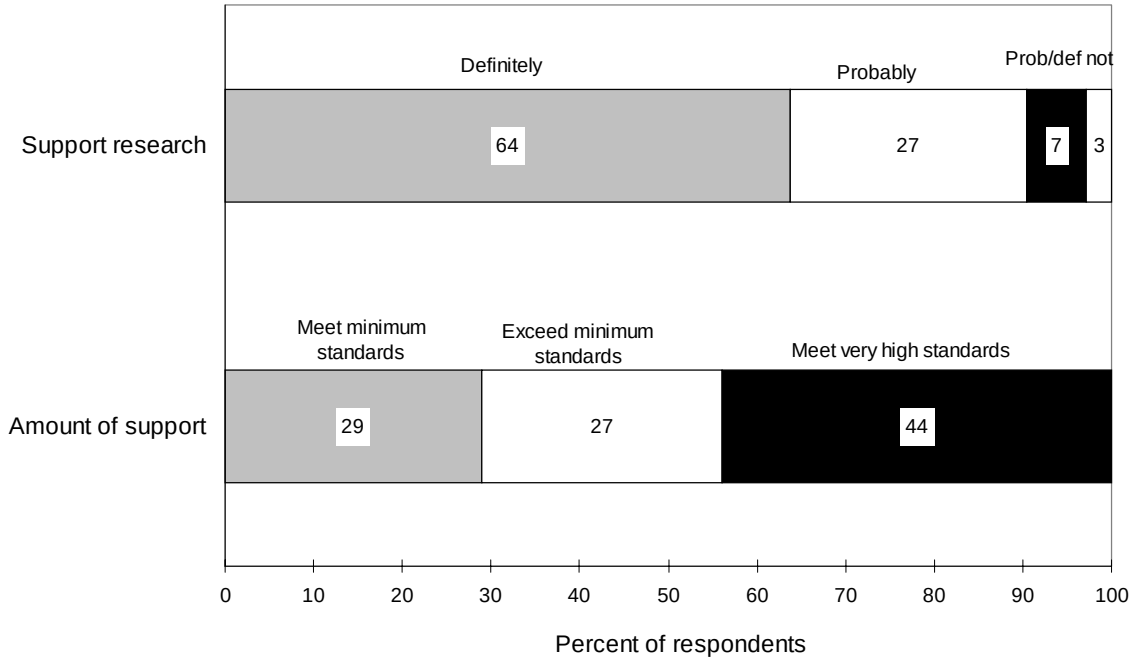
The United States participates in international assessments (such as the Third International Math and Science Study [TIMSS]). Since NAEP also assesses performance in mathematics and science, the National Center for Education Statistics (NCES) is involved in efforts to link NAEP and TIMSS scores. This linkage would allow NAEP scores for states to be compared with those of different countries.

Respondents were asked about the value of this kind of research and the level to which this kind of research should be supported. Nearly all (99 percent) of the

respondents had an opinion about the government supporting this kind of research. These respondents were overwhelmingly in favor ($p < .001$) of this kind of study: about two-thirds (64 percent) were definitely in favor of supporting it and another one-quarter (27 percent) felt the government should probably support this kind of research. Only 3 percent of the respondents felt the government should definitely not support this kind of research (see figure 2-18). Additionally, 71 percent of the respondents felt this effort should be supported to meet more than minimal standards of validity and reliability -- in spite of the increased costs that would be associated with such efforts.

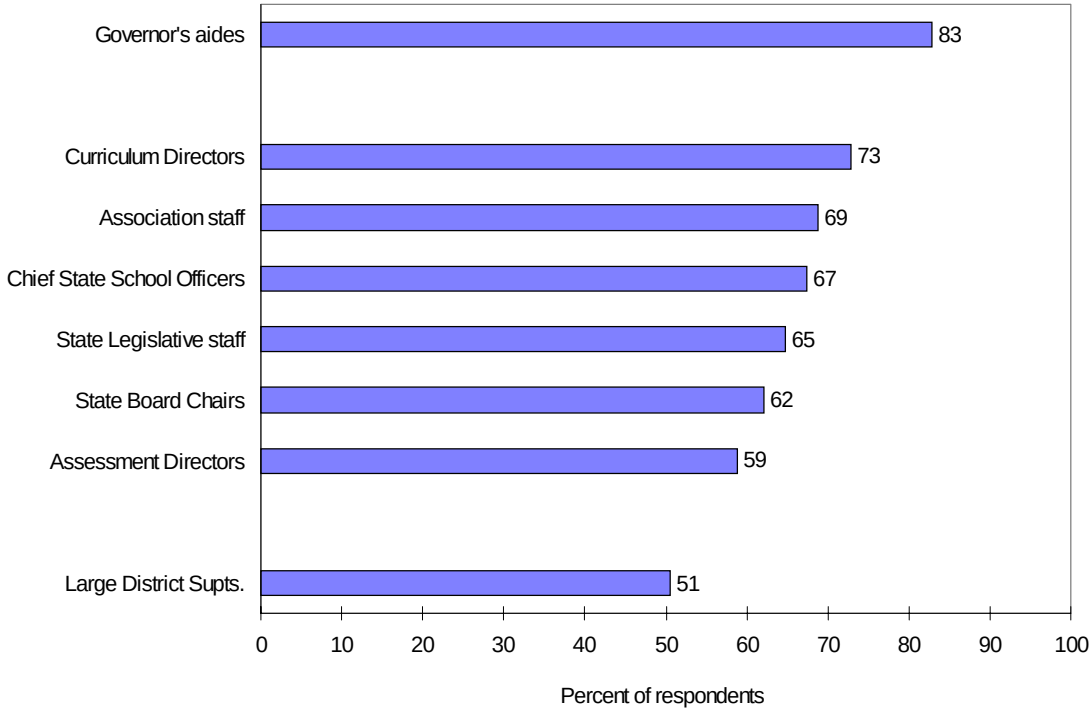
Arguments in favor of supporting this research were reflected in comments on the fact that the U.S. is part of a global economy. Accordingly, there is a need to compare ourselves with other nations as well as other states.

Figure 2-18. Linking NAEP to International Assessments:
A. Should the government support this kind of research?
B. At what level should government support this research?



Superintendents of large school districts, as a group, were the weakest supporters of this kind of research. Proportionally fewer superintendents than Governors' Educational Aides would definitely support this kind of research ($p < .05$, see figure 2-19). Still, about half (51 percent) of the superintendents were strong supporters of efforts to link NAEP with international assessments.

Figure 2-19. Supporting NAEP-International Assessment Linkages: Percentage of respondents saying "definitely support"



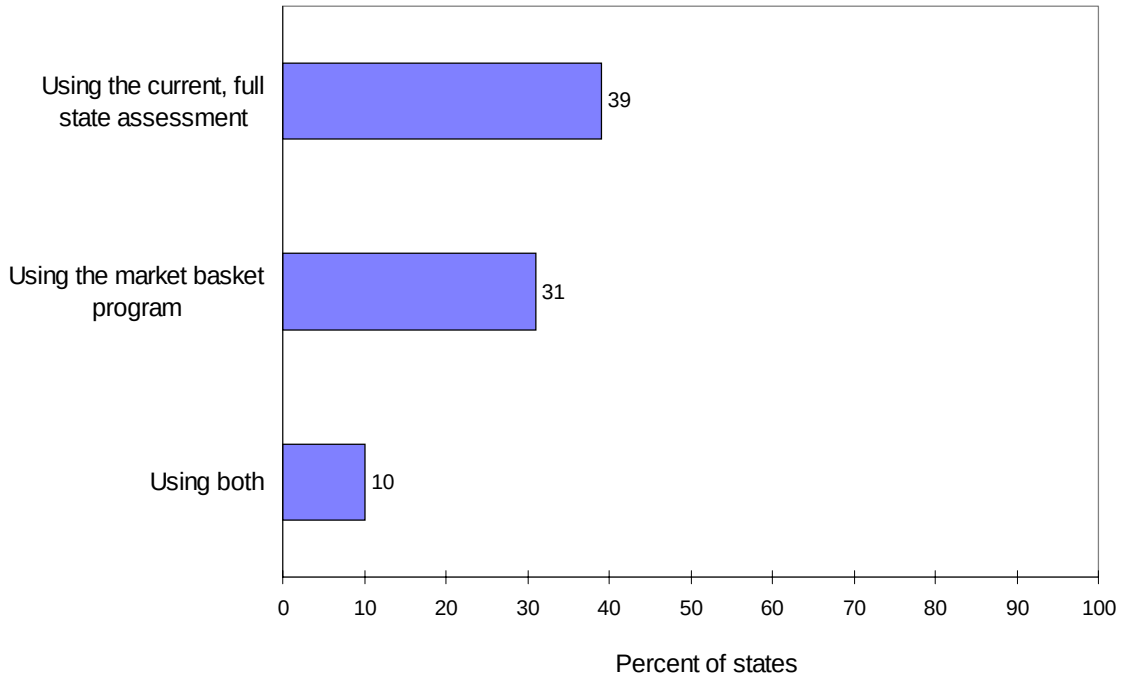
2.12 Obtaining State-Level Results

The current procedure for producing state-level results consists of administering the regular national assessment to a representative sample of students, specifically chosen for this purpose, in each participating state. This enables a great deal of information to be provided, but it is time-consuming to implement. As an alternative, a "market basket" method is being considered. This approach would provide representative sets of assessment exercises to participating states. They could be used to obtain state-level results, to calibrate the state's assessment with NAEP, or to obtain state-level results on a more frequent schedule than the current NAEP. Comparable proportions of states (39 percent and 31 percent, respectively) indicated they were definitely interested in the

current state assessment and market basket program. Only 10 percent indicated they were definitely interested in both (see figure 2-20).

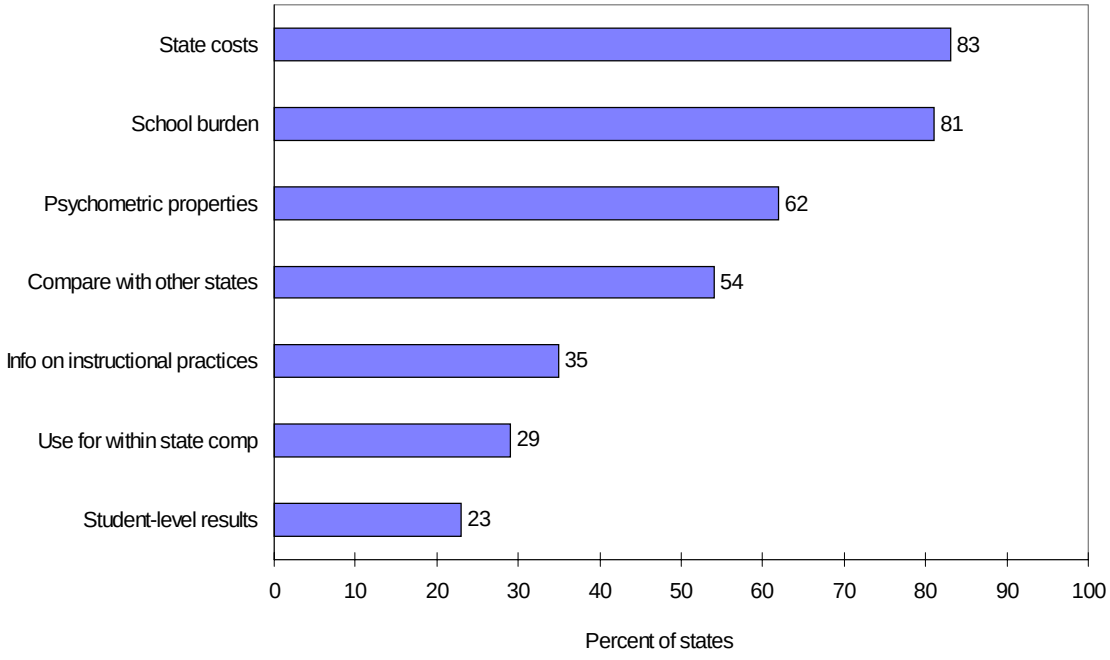
Very few respondents saw a need to comment on using the current assessment. The market basket concept evoked interest from several respondents. However, there was a concern that more information was needed to permit a better evaluation of this alternative.

Figure 2-20. Interest in Different Approaches for Obtaining State NAEP Results: Percentage of states "definitely interested"



In order to decide among alternative NAEP approaches, certain factors must be considered. The relative importance of different factors was assessed by asking how important each of seven different factors was in evaluating assessment approaches. The most important factors, evaluated as being "Very important" in at least 80 percent of the states, were state costs and school burden. These two factors were clearly more important than any other factors (except for psychometric test properties, $p < .01$, see figure 2-21).

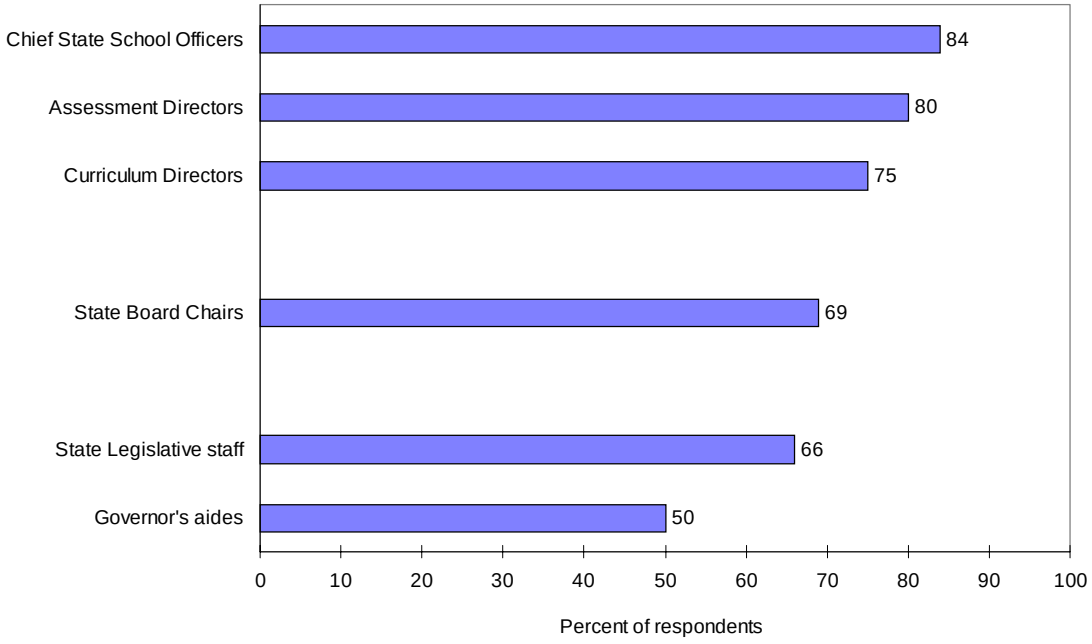
Figure 2-21. Criteria for Evaluating Alternative Approaches for Obtaining State NAEP Results: Percentage of states rating criterion as "very important"



Even though NAEP is not intended to provide individual student-level results, nearly one-quarter (23 percent) of the states felt this was a very important factor in their considerations of assessment approaches. One respondent indicated that this was the reason why his state did not participate in NAEP.

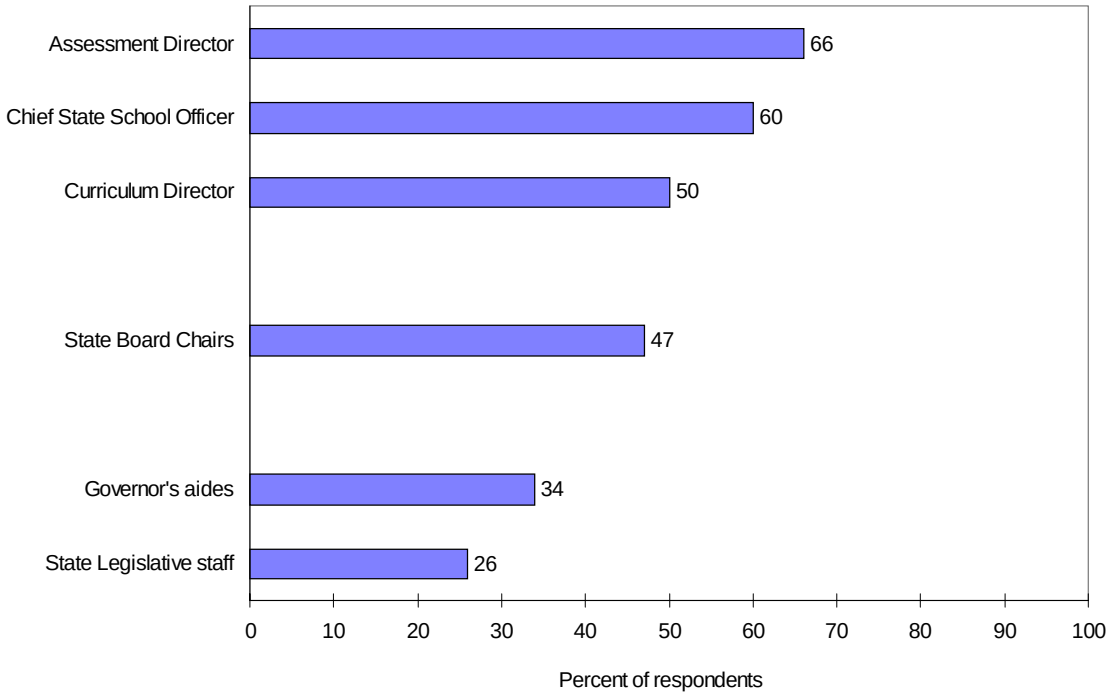
Two of the evaluation factors differed in importance among the different constituencies that were surveyed. State costs were very important to the State Education Agency representatives. At least three-quarters of the Assessment Directors, Curriculum Directors, and Chief State School Officers felt costs were "very important." In comparison with the nearly five-sixths (84 percent) of the Chief State School Officers who felt costs were very important, only half of the Governor's Educational Aides felt this factor was very important ($p < .05$, see figure 2-22).

Figure 2-22. Importance of State Costs: Percentage of respondents saying "very important"



Similarly, psychometric test properties were more important to Assessment Directors than they were to State Legislative staff ($p < .01$) or to Governors' Aides ($p < .05$, see figure 2-23). These political constituencies probably recognize the limits of their expertise and may be taking the simple perspective of "if we can't understand it, the public can't understand it, so it can't be too important."

Figure 2-23. Importance of Psychometric Test Properties: Percent of respondents saying "very important"

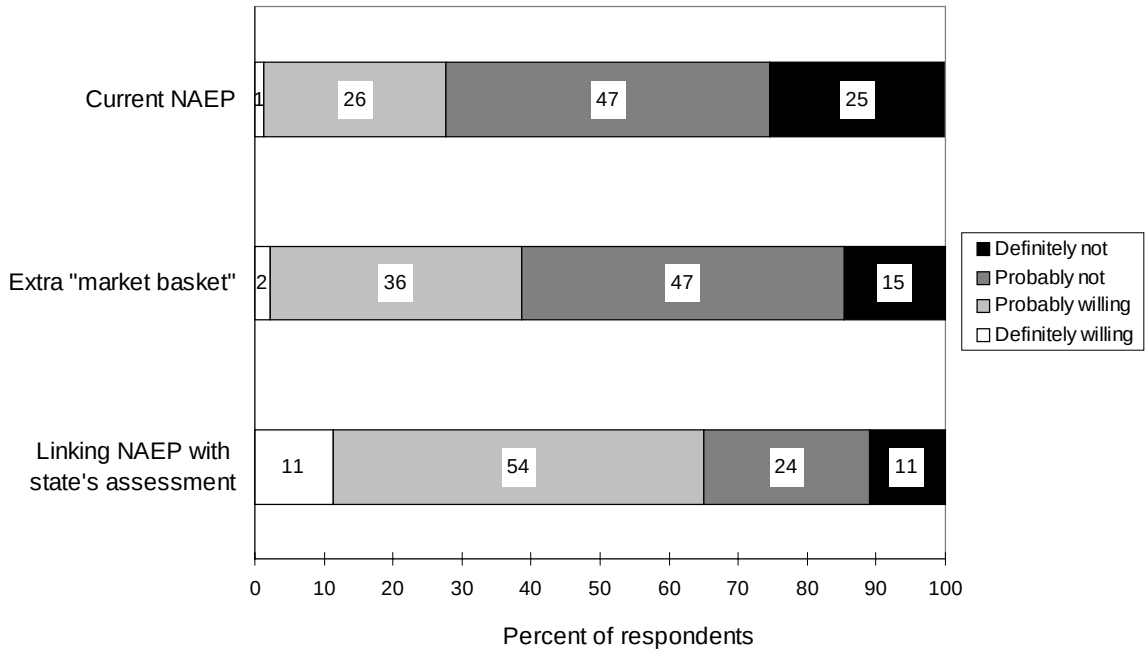


2.13 States Paying for Some Services

In order to identify the services that will be of greatest value to participating states and to explore different structures for finding services in the future, respondents were asked to assess their state's willingness to pay for three different services (a state-level assessment using the current NAEP approach, linking NAEP results with the state's regular assessment, and the provision of extra "market basket" assessments for the states to use as they desire).

The most popular of these services was linkages with the state's regular assessment program ($p < .001$). About two-thirds (65 percent) of the respondents indicated that their state would probably or definitely be willing to pay for this service. The next most popular of these were the extra market basket assessments ($p < .01$). Only about one-quarter (27 percent) of the respondents indicated a willingness to pay for a full state-level assessment using the current NAEP approach (see figure 2-24).

Figure 2-24. States' Willingness to Pay for Some Services: Percentage of respondents indicating level of "willingness"



3.0 FOCUS GROUP RESULTS

The focus group protocols were designed to gather information similar to the data which were collected on the NAEP Constituents' Surveys. This section summarizes the results of the sessions. It also summarizes the responses and comments from the AERA and NCME officers who completed and returned surveys. The comments in this section represent the feelings of the three to ten members of each constituency who attended the focus groups. They are not necessarily representative of the populations of the seven constituencies of which they are members. Comments cannot be generalized beyond the group members that participated. Nonetheless, their insights provide useful feedback about NAEP.

3.1 Background Information - Emphasis

Public secondary school principals who participated in the focus group felt questions about *school characteristics* were important to include in NAEP. *Instructional practice* data are useful, but detailed, carefully worded questions are necessary in order to get valid and useful information because instructional practice terms vary by grade level. For instance, participants believed teachers at eighth grade may have a different interpretation of "team teaching" than teachers at the fourth and twelfth grade levels. Background information was felt to be important to the participants for explaining performance in urban schools at different grade levels. They also felt *student and home background characteristics* were important characteristics to study.

The private secondary school principal focus group felt it was most important to know about economic factors, family composition, family education level, homework, the climate of the homework setting, and any factors that can be collected which may relate to raising the success level of students. *Instructional practices* were interesting to them, but reports must give concrete information.

Elementary school principals at the focus group felt the usefulness of background information varies greatly by user. All of the sets of background information were important to them, and as many should be measured as possible. If priorities need to be set, participants felt *student and home background characteristics* were most important, since they are the chief indicators of student performance. *School characteristics* were of secondary importance to the group, followed by *instructional practices* and *topics of current educational relevance*. These elementary school principals were very interested in how parents support education in the home.

Members of the media writers' focus group felt that NAEP should focus its background variable collection on information that is not already well known. They felt

new information on *instructional practices*, such as teaching techniques, would be helpful because readers do not know the relationships between such variables and achievement. However, one of the group's concerns was the fact that generalizations reflect poorly on some subgroups. In terms of collecting information on current issues of educational relevance, participants believed such items should be studied only when a certain percentage of schools are implementing the new techniques. By waiting until the techniques are more common, NAEP could better detect whether changes were significant. Also, participants mentioned that since other agencies might conduct studies in these areas (e.g. the Office of Education Research and Improvement [OERI] might sponsor studies of charter schools), it may be better to leave data collection to those agencies.

Members of the public who participated in the focus group felt the most important background characteristics to collect in conjunction with NAEP were family characteristics such as income, household location, and parental education. Also mentioned was access to educational resources in the home. Race and gender were perceived to be of lesser importance. Participants mentioned that knowing the outcomes of one group of students versus another might help schools know "where to focus" their efforts on educational improvements. A discussion of the lack of discipline in the classroom and the link with teacher qualifications and teaching skill ensued. Some participants thought schools should use NAEP data to find out what types of disciplinary techniques were used 25 years ago and return to them. In addition to teaching style, participants mentioned class and school size, type of school, and educational resources available in the school as being important to measure.

The focus group participants from the business-in-education programs expressed that they would be interested in data on *school characteristics*, including community involvement of outside volunteers, senior citizens, businesses, and parents. They had never seen *instructional practices* used as factors, but felt that differences in teaching styles would be useful. In terms of *student characteristics*, social and economic status should be explored, possibly instead of race/ethnicity since such categories are becoming more difficult to define. They emphasized the importance of site visits and student reporting as data collection methods. The group felt there was a saturation of self-reporting, and suggested that video observations might be a more useful tool to gather background information on classroom teaching.

The teacher focus group expressed the desire to have information on spending per pupil, class size, school building conditions, mobility, ESOL participation, and at-risk populations. Participants also felt background information should be collected on licensing of teachers. The group stated that information should be collected on the use of textbooks and computers, and on the use of business/school links. Additionally, they

were interested in finding out what states use high stakes assessments, and the amount of planning time and collaboration that teachers have available to them. These teachers stressed that such breakdowns are critical.

The majority of education researchers surveyed felt that many of the background characteristics were "very important." About 83 percent felt *student factors* were "very important", followed by *school characteristics* (70 percent) and *instructional practices* (60 percent). When asked about the importance of *topics of educational relevance*, respondents felt such information was "somewhat" to "moderately important" (37 percent for each category). Some respondents noted their negative feelings about the collection of background variables, stating that there is much intra-school variance on such factors as well as variance between schools within a state. They stated that NAEP should focus on macro-level variables which it measures well. Other respondents, however, pointed out that such data collection is critical for the purpose of assessment and guiding educational practice. One researcher felt that it was important to include as much contextual information as possible to help understand or explain observed differences in achievement and to take well-informed action.

3.2 Background Questions - Impacts on Release of Results

Members of the public high school principals focus group felt the current delay of six months in reporting is satisfactory. Private high school principals who participated suggested that results should be released before the next school year starts. The private school principals felt it would be best if results were released in March or April, so that curriculum changes could occur. They felt a six month delay was too long generally. The elementary school principal focus group felt it would be acceptable if background/ achievement relationships were released separately from the initial reports, so that the initial reports would not be delayed. They would prefer a multi-step release of separate reports to a delay in any reporting.

The education writers who participated expressed mixed opinions to the option of removing background variables in order to shorten the length of time before reporting. Some writers felt that if reports could come out sooner, maybe NAEP shouldn't collect background data. Others pointed out that two separate reports would work well: an initial report without background relationships, followed by a more comprehensive report at a later date. One member expressed interest in such a breakdown, and felt it was more critical than the timing of the report. In terms of a reasonable delay, the group noted that if results could be reported in the same year as the testing, it would more likely be published and would probably double the press exposure.

Members of the business organization group felt that fewer background variables might be used, so that reports could be released at an earlier time. All three participants felt it might be preferable to have a narrower range of factors that are explored in more depth.

Teacher representatives who participated felt that data should be released as soon as possible. Participants noted that NAEP is influential and provides useful information which educators can use to provide rationales for their instructional practices. At the state level, an early release is important for curriculum decisions. Background information is relevant to the participants and is worth a reasonable delay of three to four months.

Fifty-seven percent of the education researchers indicated that all background information should be included and that a short delay is reasonable. They felt the background information is essential in the interpretation of NAEP results and that student performance is accounted for largely by background factors. One researcher pointed out the importance of understanding relationships between specific demographic information and initiatives and school outcomes. Several respondents suggested a multi-stage release, with achievement data released first, followed by a full report at a later date. Those who suggested that no background variables be reported (7 percent) felt that the information is only useful when collected and reported at a local level. They also believed that the media and public pay little attention to the breakdown of results by background variables.

3.3 Technical Documentation of NAEP vs. Timeliness of Reports

Public high school principal participants agreed that the current delay in reporting, to allow for the inclusion of technical documentation, was satisfactory. Good documentation is very important, so it makes sense to delay release of results for six months. The private high school principals focus group had difficulty with the topic, and no comments were made. As with background information, the elementary school principals who participated felt a multi-step approach should be used with an initial release followed by technical documentation in a later report, assuming the technical documentation would not change the reported results.

The media group suggested that technical information could be placed on the World Wide Web and that an executive summary would be sufficient for most writers. As long as NAEP feels the information is acceptable, participants felt the technical documentation wouldn't need to be published before reporting. Again, the group suggested a two step reporting process of an executive summary followed by complete documentation at a later time.

The business group had no opinion on this issue, but pointed out that anomalies with other reports would be challenged, so if some of the information was controversial, it would be a problem to report it without technical documentation.

The teachers' focus group expressed the opinion that accurate information is critical, and that the original report that comes out has the most impact. Therefore, they felt NAEP cannot afford to compromise and release the report early without full technical documentation.

About three-quarters (77 percent) of the education researchers felt that reports should not be delayed due to the reporting of technical documentation. One researcher indicated that NAEP technical reports have been outstanding and are a model for the entire field. Almost all of those who included comments suggested that a multi-stage reporting process could occur, with an initial report followed by the full report which included technical documentation. A suggestion was made to put such information on the Web until the print version was available.

3.4 Including a Parent Survey

Several principals in the group felt that NCES should try to get information from parents, since they play important roles in the education process. Public secondary school principal participants suggested that the PTA and booster organizations could be used to help encourage participation in a parent survey. Some principals mentioned that they have conservative groups in their communities who would raise objections. The private secondary school principal focus group felt that the survey should focus on home environment topics. However, questions were raised by participants as to how this would be collected (by the school or NAEP). These principals said the response rate is usually very small on parent mail surveys, and only provides a good representation of parents who care about the topic. Many elementary school principals in the group felt a parent survey should be used to assess parental involvement in education practices at home. However, the packaging is critical, since the survey must not appear to be an invasion of privacy. Participants noted that the cover letter should state that participation is voluntary, and that parents will not be identified by the government. It should state what the data will be used for in advance and should offer to send parents free information after the results are reported. Positive marketing is crucial, perhaps through meetings, and surveys could be disseminated during school functions. A few principals believed that only parents who are less fearful and more involved in achievement issues might respond, so this should be noted in any reports which include parent survey data.

Participants from the media were skeptical regarding the possibility of collecting representative data from parents, and expected only a two to three percent response rate.

They felt that telephone interviewing would be necessary. Although they thought parents would be the best source of many types of information, some expressed a need to stick to measuring the basics, and others suggested that NAEP may not be the best vehicle to collect such data. Suggestions were made that NAEP coordinate with another agency which collects data from parents, or that the information be collected only for descriptive purposes.

The general public focus group thought that it would be important to measure the level of parental participation in school and to obtain the parents' perspective on school performance. However, some participants mentioned that county schools already do this, and could not imagine how information collected on a larger scale would be useful. Others mentioned that parents are too busy to be involved in their children's schools and that education suffers as a result. The participants generally agreed that parents' educational expectations for their children are crucial, and no one thought that a parent survey would be too intrusive or burdensome. Again, however, national findings about parental expectations were thought to be "too big, overwhelming." One participant raised the issue of having to translate the parent survey into many languages, and others thought parents might not be honest. Everyone wanted to know how data collected from parents would be used to improve education. They expressed reservations that a lot of data would be compiled but not used.

The business focus group felt that the collection of information from parents would be a political nightmare, but that it might increase parent involvement in education. However, one member pointed out that parents might not feel the need to respond, since they don't receive any information on their child's performance.

Members of the teachers' group noted that some parents may appreciate the survey because parents are interested in providing background information on their child's home activities. However, there are several potential problems, because some parents can't read, and others may not be truthful about their home practices. Participants felt NAEP would need to explain the rationale for data collection to alleviate some of the feelings of intrusiveness. These teachers were interested in the relationship between performance and television viewing time or parental education. One teacher was also interested in comparing the performance of "latch-key kids" with other children.

The educational researchers who were surveyed did not respond with a consensus or voice strong opinions on the use of a parent survey. Forty percent felt that NAEP should "probably have a parent survey", while 33 percent felt it probably should not. Only 27 percent of the researchers felt strongly about the inclusion of a parent survey, with 13 percent strongly in favor and 13 percent strongly opposed to the concept. Positive comments from researchers stated that such information would enhance the

richness of NAEP data, and that the gain might be worth the potential risk. Those who were not in favor of the parent survey mentioned that it might stir up trouble and would lead to political concerns, great costs, and burden. They expressed concern that low response rates would bias the sample. Some respondents suggested that the parent survey be pilot tested to judge its usefulness.

3.5 Reporting: Schedule for Release of Results

All principals in the focus groups felt it would be more helpful if reports were released on a set schedule each year. There were mixed opinions as to whether the results needed to be released all in one week, or could just follow a set release schedule throughout the year. Participants believed a set schedule would allow principals to know when the information will be available and to know whether the information should have arrived. One principal commented that, "we as principals live by schedules and would like to be on a set schedule." The group members noted that schools like to prepare before such information is reported to the public, so they can inform parents about the sampling process and other issues.

Education writer participants all agreed that a set schedule for report release was preferable. They would appreciate receiving the anticipated schedule in advance to help them plan their reporting space. One writer mentioned that if the data are all released in one week, the press will not have enough space available to report all of the results.

The general public focus group said that data should be reported more quickly and publicized widely. They said that releasing all of the findings from an assessment at one time would improve the attention paid to NAEP. They asserted that they would read about it if it were published in a newspaper, but also thought television was an appropriate medium for releasing the data. Parents in the group thought their schools should inform them about the results of the assessments.

The business organization group felt that a set schedule would help people to start thinking about NAEP before the reports are released. They suggested that NAEP reports hook on to the Goals Panel Report, which explains the results, resulting in a sharper "State of Education" report. The three members offered different vehicles for reporting, including electronic town meetings and summits planned around the releases.

All members of the teachers focus group stated that a set release was preferable, since they like to know when to expect reports. They felt it is important that the data be presented clearly to the public at the release time to facilitate understanding.

The majority of responding researchers (57 percent) preferred to have NAEP releases occur as they were available rather than releasing on a set schedule because they like to have results available as soon as possible. Those who did prefer a set schedule (23 percent) suggested that it would be helpful for comparisons of various subjects and disciplines if all results were released at the same time. One-fifth (20 percent) of the researchers did not have a preference about the release schedule.

3.6 Assessing Subject Areas in Combinations or Individually

Most of the participating principals agreed that the specificity of individual subjects was necessary to relate achievement to background information. In the areas of social studies and history, these principals felt a composite score would not be needed, but rather detailed information on the subjects. The level of specificity depends on the grade level. Differentiating these subjects at the elementary level was not important to the participants. More principals in the group noted the importance of assessing the subjects individually rather than as a cluster in the social sciences and history. In science, most principals who participated felt that the three areas should be assessed separately. A few indicated that the science areas could be clustered, especially in the fourth grade. In language arts, all principals in the group asserted that reading and writing should be assessed and reported separately.

Principals who participated also suggested some additional areas which should be assessed in national testing. Foreign language was suggested, since they viewed it as a necessary skill for students. They also felt a grammar/language usage component of the writing assessment was needed.

The media group expressed many different opinions on the issue of clustering subjects versus assessing them in isolation. One member pointed out that reading and writing were not isolated subjects, and another countered that they were never clustered in assessment. Clustering made sense to the group, because subjects could be intertwined. However, they stressed that assessing in isolation could help guide decisions better than a cluster assessment.

The business group initially had no opinion; however, they did note that most states do not have economics or civics assessments. They suggested that the only areas to be assessed individually should be those that are assessed at the state level (e.g., reading and writing). Participants in the group would always like to have as much specificity as possible. In addition, they felt that reading and writing were usually not integrated in the curriculum.

In social sciences and history, the teachers focus group preferred assessing subjects more frequently, and they felt geography and history should be combined at all grade levels. They felt civics, economics, and art should not be assessed before eighth grade. For science, all group members felt that the subjects should be clustered for fourth and eighth, and then assessed in isolation for twelfth grade. The group indicated that reading and language arts should be assessed separately.

Participating researchers were split on whether social science and history should be assessed individually (42 percent) or as a cluster (52 percent), with 6 percent having no opinion. They were also split in the area of science, with 45 percent favoring assessment in isolation and 55 percent preferring cluster assessment. There was a stronger consensus in reading and writing, with 87 percent favoring individual assessment. Some researchers commented that subject-specific data are imperative, particularly if they allow for linkage to background information. Another researcher pointed out that the integration of curriculum should be encouraged.

3.7 Desire for Information on Skill Areas within Subjects

The principals who participated felt it was important to include as much specificity in reporting as possible. Without subscale breakdowns, they believed information could not impact curriculum changes. This information is related closely enough to state curriculum breakdowns to be relevant and useful. Subscales provide a meaning to the overall scores and allow a benchmark for progress. Science was especially important to the group; however, science, geography, and history are not as easy to break down at the fourth grade level. Some members noted that information on skill areas promotes targeting of efforts to areas of deficiency and allows for tracking of progress in a way that is consistent across states or districts, which is important, given the extent to which students move. Participants felt the writing subscales are better formulated than the other subject subscales, because they deal with intellectual processes, not just content. Some of the high school principals were concerned that the subscale scores are not accurate due to the lack of motivation on the part of students, teachers, and schools. They also felt the media does not generally use subscale scores in reporting.

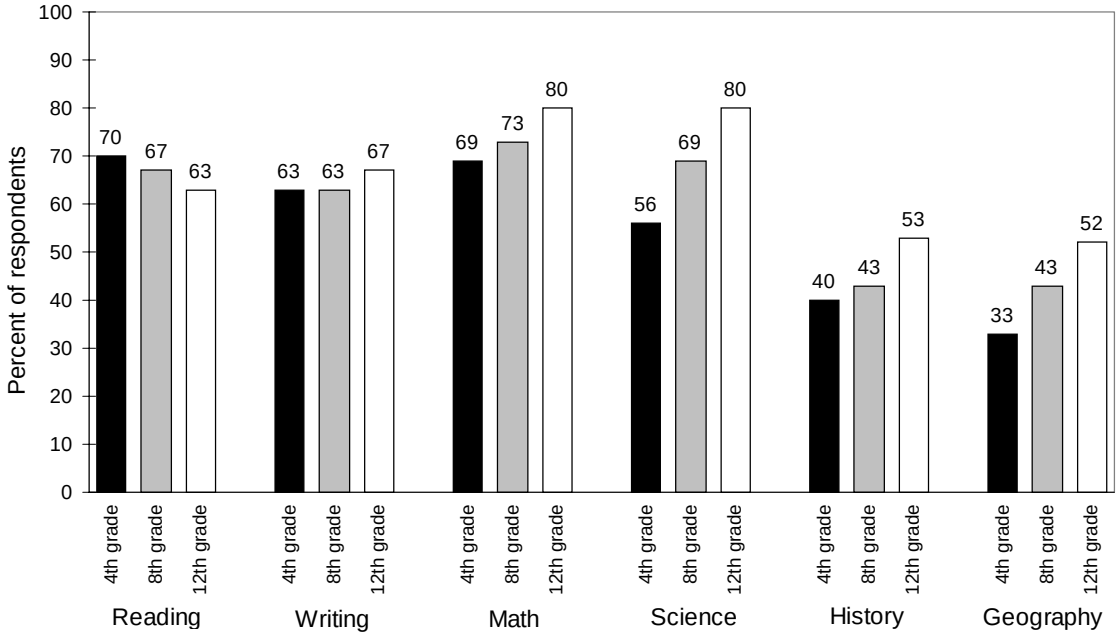
The education writers who participated believed that the more detail provided on test scores, the better. With more detail, writers can identify where weaknesses occur and could see what children are and are not learning. If subscales are left out, scores are massed together. One member stressed the need to assess subject areas in ways that do not also measure reading, since some students may have learning disabilities which could interfere with their scores. The group felt subscale scores were important for fourth grade and that reading subscale scores were critical at all grade levels. They also noted that subject area breakdowns in geography made no sense to them.

The business group stressed that detail within categories was very important. They noted that subject area breakdowns need to be specific and in topic areas that are meaningful to people. Otherwise, communication gaps might occur, because the public and reporters won't know what the scores indicate. The three participants agreed that more detailed information on skill areas within subjects was needed, such as a description of what the subscales mean or a further breakdown of each subscale. They indicated the need for subscales was universal across all grade levels to identify any curriculum problems as early as possible. This group mentioned that there might be a need for more basics at the fourth grade level, such as grammar. Other subjects they suggested for national assessments included technology, listening comprehension, teamwork, and SCANS skills; however, one member felt these would be difficult to assess at a national level, due to political problems. She suggested that assessment of these skills could be integrated with assessments of existing subject areas.

The teachers' focus group stressed that the basics were the essential areas to assess. In fourth grade, they believed subscale scores in reading, writing, and math were critical. In all grades, they felt subscales would be valuable if done correctly. Participants felt breakdowns within subject areas would help the public understand what is included in the test and also explain what the state curriculum covers and what it does not yet cover. Subscale scores for states would be especially useful to them to assess how students are doing, and results could be used to evaluate teacher requirements and shape teaching curriculum at universities. One member suggested that technology should be assessed so that it would drive all states to provide improvements in their schools. Another member stressed the need for national assessments in the core academics.

With the exception of geography and history, over half of the researchers felt that subscales were "moderately" or "very important" at all grade levels (see figure 3-1). In reading, researchers felt breakdowns were "moderately" or "very important" at the lower grade levels. About two-thirds of the respondents noted that writing subscale scores were "moderately" or "very important" for all grades. In math and science, the number of researchers indicating that subject subscales were "moderately" or "very important" increased at the higher grade levels. Similar patterns appeared for geography and history, although fewer researchers felt that subscale scores in these two areas were "moderately" or "very important".

Figure 3-1. Desire for Subscale Information, by Subject and Grade: Percentage of respondents saying information is moderately or very important



3.8 Frequency of Data Collections

The principals attending the focus groups did not favor changing the schedule to decrease the number of students tested at any school in a given year, because they believed it would increase overall burden. Their concern was the number of times their school had to participate in testing rather than the number of students tested at each time. They felt the current sample already disrupts the classroom setting when small groups of students are pulled for assessment. A suggestion was made that NAEP should test the entire class instead of a sample to provide less classroom interruption. These principals felt that if yearly testing does occur, they would prefer to have subjects assessed more frequently versus assessing more subject areas. It would be difficult for group members to assess the impact of curriculum changes from one assessment cycle to the next if assessments are administered more than a few years apart.

This topic was not discussed with the media group, education writers' group, or the general public group. The business group participants had no opinion and said that they cared more about receiving consistent reports than the frequency of data collection.

The teachers' focus group preferred to measure subjects more frequently as opposed to assessing more subjects or spreading the burden across years. They felt that it was important not to spread the testing too thin at the fourth grade level, but to stay with assessment of reading, writing, math, and history. Participants noted that students already take so many tests that additional tests would not desirable.

About two thirds (63 percent) of researchers surveyed preferred collecting NAEP data annually; the other 37 percent preferred to maintain the current schedule, pointing out that the two year cycle imposes less burden on schools. Of those who favored yearly assessments, a consensus was not reached as to the rationale for yearly testing. Comparable proportions favored yearly testing because they felt it would reduce individual school burden, permit assessment to take place more frequently or in more subjects, or would both reduce school burden and allow for more frequent assessment or assessment in more subjects.

3.9 Linking NAEP to International Assessments

All of the principal focus groups felt that the linkage of NAEP to international assessments should be made. Much concern was voiced regarding the validity and fairness of the international comparisons themselves. Participants agreed that comparisons must be made on parallel measures. Tests must measure the same topics on students of the same age, and during the same time period. One principal stated that, "we need to examine what we are looking for, since cultures and expectations differ." Principals in the sessions indicated that NAEP needs to be redesigned to provide fair and valid comparisons with other countries. Participants felt that states and schools need as many comparisons as possible, so this information would be useful.

In terms of government spending on international comparisons, the principal focus group participants felt that the government needs to support analyses adequately to make quality comparisons. They felt quality and effective comparisons were critical, and the government should spend as much as necessary to do it right. However, one member pointed out that if the difference in cost results only in an increase from a 95 percent to a 99 percent confidence interval, the extra expense would not be worthwhile.

The media group felt that international comparisons were definitely useful and indicated that the government should collect and assess such information as well as they could. They stressed the importance of focusing on both private and public schools in such linking.

The business focus group also thought a link to international assessments was very important. They noted that businesses use such information, in addition to national data, when basing decisions on where to locate their headquarters. It would be interesting for them to see how the top scorers in a state compared with the country as a whole and also other nations. The participants felt this information would have policy implications, since it would show where the education system was and was not performing well. The group felt it was important that results be clearly explained and understandable, and that data collection and linking be done as well as possible.

The teachers' group felt that NAEP should definitely be linked to international assessments to take advantage of the available resources, as long as both assessments are accurate. One member pointed out that students are members of a global society and are competing with students in other nations. In terms of governmental support, teachers in the group felt all states should fully support such research. One member questioned whether NAEP should be the one supporting the international link if other assessments are already making such a linkage. If NAEP is the sole assessment doing so, then full support should be given.

Forty-five percent of researchers surveyed asserted that the government should definitely support research in international linkages to NAEP. However, when asked about the amount of monetary support that should be directed in this area, the respondents were split in their opinions. The majority of respondents (64 percent) noted that an international linkage is a vital area of national concern, and should be given strong support. Many pointed out, "if you do it, do it well." The minority (36 percent) who felt that the government probably or definitely should not support international linking expressed that they would rather see the resources directed to U.S. education, and pointed out that there would be differences in the test structure and sample across countries.

3.10 Impacts of Results (General)

High school principals in the focus groups generally hear from parents who care and are interested. They felt there was no impact of state results in the community when math results were released recently. The participants felt that legislators used national statistics to feed political biases. One member stated that when reports are released, the information goes to many individuals who don't know anything about NAEP. Principals who participated pointed out that the public thinks all students were tested -- especially since it is called the Nation's Report Card -- and the media impact caused a snowball effect. They felt assumptions were made when the sampling procedures were not included in reports, such as the incorrect assumption that only public school students were assessed. Some expressed concern that the reporting of scores added to the public's general disenchantment with schools. Reporting test scores close to budget decision time

could be politically harmful. Some of the group felt strongly that results were not reported positively. A suggestion was made that such results needed to be countered with positive information, such as improvement in scores over time.

The principal participants also believed NAEP results could benefit principals when information was given that related background variables with achievement. Anything that could support existing programs or encourage further parental support would be welcomed. They noted that schools would like to be able to show what was working, such as special programs for single parent families. Reporting could be beneficial if it was used to stress the importance of parental support for education and the sharing of responsibility between home and school. It was mentioned that the specificity of reports can be very useful in making inferences about programs and techniques, whereas general scores would not helpful. If principals of participating schools could get an executive summary or brief report in advance of the public, it would help them to provide explanations and hold informational sessions. The participating principals felt that the results of NAEP could be used to shape curriculum in teacher preparation courses at universities. A schedule of assessments and report releases, in addition to copies of all NAEP reports, should be available on the Internet. Also, one principal stated that NAEP test curriculum should be tied to the National Standards in each subject area.

Members of the media group pointed out that they were interested in reporting on NAEP to the extent that they could get good information. The reports need to be easy to understand, especially since there is a high turnover rate for reporters of local papers. One participant noted that NAEP was hard to sell with results being reported on state assessment programs, so television might be a better source for reporting. Also, NAEP reports usually do not make the front page. The group felt state NAEP results were more reportable, since education writers have limited space available. In addition, these writers usually compare recent scores to scores from previous years. They would like to be able to report information without numbers and graphs, while noting that better graphics would be more helpful to them.

The members of the general public attending the focus group said that they would like to know whether "we are improving as a nation" educationally. They expressed concerns that we are not competitive with other nations. They also mentioned that basic skills that were not likely to be influenced by regional curricula should be measured on a national level. Participants approved of compiling data by state so that residents could compare their state's scores to national data to see how they ranked. Participants named other statistics that they thought they might want to know, such as number of graduates and number of dropouts, but they were unclear as to what value those findings would have for them. Focus group members said that NAEP results should focus on how to improve schools. In other words, they would be interested in knowing not just how

students in their state ranked according to national standards, but also what had been done in the highest ranking states and what was going to be done to improve their performance based on the findings. A discussion centered on whether members of the general public who were not parents of school children would be very interested in educational data or educational improvement. Opinions varied, but a clear majority of these focus group attendees agreed on two issues in the discussion: no one wants to pay higher taxes even to improve education, and any reform must take place at the local level. Participants perceived "the system" at fault and were well informed about aspects of it that could be improved (e.g., more resources, more emphasis on basic skills, better trained teachers, release of inadequate teachers, more discipline in school, more parent involvement), but they did not seem to see any role for the community (the general public) in reform.

The three members of the business focus group felt that the business community is not very familiar with NAEP results. However, they noted that the governors' offices use SAT and NAEP scores in economic development to make comparisons and to compete for money. One member in the group stated that companies consider relocation sites based on state assessment scores. However, all the participants believed that few local educators pay attention to NAEP scores since they are not linked to their state assessments.

The teachers' focus group pointed out that teachers were not as aware of NAEP scores as they were of state and local scores. One participant felt teachers tended to get defensive about results due to the negative impact of low scores. Participants agreed that assessment results were now watched more closely than they were in the past. They believed that the media generally put a negative spin on headlines even when there were many positive results. However, a few teachers in the group pointed out that NAEP can be used to provide an honest presentation of what is taking place in schools and to improve and to analyze what school systems need to do. It can also be used to celebrate what schools do well, identifying strengths. It can also identify weaknesses, marking areas for improvement. The group thought background information helped to clarify the results and provided a more accurate story.

A few of the researchers who completed the survey pointed out NAEP was not designed to produce scores for districts, schools, or students. The participating researchers preferred to keep NAEP as a low stakes assessment. They feared changes could weaken or destroy comparability with previous data. One researcher felt that NAEP contains useful state-level data which could be used by researchers to focus public attention on problem areas, but it should not become another measuring stick for individual district comparisons. Another researcher stated that NAEP should realign itself to reflect performance based assessment.

3.11 General Public Focus Group Findings

The focus group with the general public included topics that were not discussed with other constituents. These are summarized below.

General Impressions of Educational Assessment. Participants focused at first on the type of individual achievement testing with which they were most familiar. They concurred that there was some value in having such testing "give the average," as one participant put it, but mentioned that the achievements of some students were poorly measured by standardized testing.

The group participants failed to see how national educational achievement could be adequately measured because of regional differences in curricula, student demographics, and school resources, particularly the quality of teachers. The notion of testing on a national level and the utility of findings from such an assessment were difficult to grasp. They also were not able to mention ways in which the findings from such assessments could be used.

Advantages of a National Assessment. Participants mentioned four reasons to conduct national assessment: 1) to prepare students for employment; 2) to measure students in the U.S. against students from other nations; 3) to assess teacher performance; and 4) to motivate parents to become more involved in their childrens' schools.

They also offered some caveats. For instance, they seemed to place importance on the international standing of U.S. students, but they pointed out that education was not valued as highly in the U.S. as in some other countries. The focus group members felt that students were not motivated to study as hard in this country as in others. This implies that U.S. students' poor performance in comparison to students from other countries may be due to reasons unrelated to curricula or teaching technique.

Disadvantages of a National Assessment. The group believed that national assessment does not take into account regional differences, resources available to schools (qualified teachers, current textbooks, etc.), and the cultural backgrounds of students, which may cause them to perform less well on standardized tests.

The general public focus group was most concerned with what would be done with the test results. Although they approved of the assessments as a way to "develop basic standards applicable throughout the country," they could not envision how the results would be used. They strongly believed that the assessments should focus on the basics. There was even a spirited discussion of the drawbacks of technological innovations in the classroom. "The real question is who is going to deal with results."

One participant stressed that results would be "useless" if given to administrators because only teachers could bring about change in education. In general, they were skeptical of national efforts to improve education and believed that any effective reform must be implemented at the local level. However, they failed to see how data from national assessments could be used at the local level.

Knowledge of the National Assessment of Educational Progress (NAEP) and the Federal Government's Role. The participants in the public focus group had never heard of NAEP. One asked if it was related to the National Education Goals. In the discussion that ensued after NAEP was briefly explained, the participants expressed their disapproval of the federal government in general and the Department of Education in particular because they "put everybody in the same big group," ignoring regional differences. Participants generally agreed that educational data should be collected at the state and local level, and some pointed out that the federal government was paying for that to be done with NAEP. Federal financing of studies was met with approval, as was the federal government's compiling national data for international comparisons. However, participants perceived that the federal government "controlled" the data collected through NAEP and disapproved of that.

Summary of Public Focus Group. Members of the public who participated in the focus group had no knowledge of NAEP and believed a national assessment of educational progress to be of limited use. They were doubtful that this type of research was an appropriate effort for the federal government. They strongly believed in local control of the schools and in intervention at the classroom level to improve education.

By the end of the discussion, their concern had shifted somewhat to focus on the fact that NAEP is poorly known. The participants were united in their belief that rankings of U.S. students or students in each state were not valuable unless they were accompanied by reasons why students rank as they do and strategies for improving education. As one participant remarked, "The data are not going to solve problems." Participants expressed interest in knowing how the data has been used since NAEP was first administered. They also wanted to know how education in the U.S. had been improved as a result of the study over the 25 years that NAEP has been conducted.

APPENDIX A: NAEP Constituents' Survey
(Version I: Long Form)

OMB #:1880-0529 Approval expires: 2/28/98

Survey of NAEP Constituents on Future Directions for NAEP

The National Assessment of Educational Progress (NAEP) must respond to a wide array of important, but diverse, purposes and interests. The National Center for Education Statistics (NCES) is supporting an effort to redesign NAEP to make it more responsive to your needs and interests. In order to do this, NCES needs your input. By answering these questions, you can have a major impact on decisions regarding the future of NAEP.

After each item, we have provided space for you to elaborate on any of your responses. Please be assured that all responses will be kept confidential: your responses will not be identified, nor will your state's or organization's responses be identified in any reports. We will report only summaries for various groups of respondents. If you have any questions or need assistance in order to respond, please call Amy Rathbun at AIR (D.C.)(202)944-5254 or Roger Levine at AIR (CA)(415)493-3550 for clarification. ***Please return this survey by March 11, 1997.***

1. BACKGROUND INFORMATION—EMPHASIS

NAEP collects background information about the characteristics of participating schools, teachers, and students, in addition to measuring student achievement. These data (which are kept confidential) are collected to allow measurement of how different types of students and schools are doing. These data can also help explore the relationships between different kinds of teaching practices and student achievement.

The background information collected and analyzed in NAEP can address a wide array of factors—instructional practices and techniques, school characteristics and programs, and student background factors. To allow their prioritization, please indicate how important you think each of the following are for NAEP to study.

	Not Important	Somewhat Important	Moderately Important	Very Important
Instructional Practices (e.g., content coverage, teaching technique)	NI	SI	MI	VI
School Characteristics (e.g., size, location, type)	NI	SI	MI	VI
Student Background Factors (e.g., race/ethnicity, gender, parental involvement)	NI	SI	MI	VI
Topics of Current Educ'l Relevance (e.g., portfolio assessment, year-round schools, charter schools)	NI	SI	MI	VI

Comments:

2. BACKGROUND QUESTIONS—IMPACTS ON RELEASE OF RESULTS

Presently, data from NAEP background items must be processed and analyzed so results can be reported. While the information is valuable, this takes time and defers the release of NAEP scores. There are also costs associated with the collection of background information. If these data were not collected, NAEP could use the funds for other purposes. Which of the following alternatives do you prefer?

Background information should not be collected, so that resources can be used for other NAEP tasks and so reports can be released sooner (about six months). ☐

Some background information should be collected, but less detailed or less frequently, so reports can be released sooner. ☐

This background information is valuable, so the current practice (not releasing reports until all information has been analyzed) should be maintained. ☐

Comments: *(We are interested in your specific experiences using these data.)*

3. TECHNICAL DOCUMENTATION OF NAEP VS. TIMELINESS OF REPORTS

NAEP attempts to adhere to very high standards of technical quality. This has included providing extensive documentation of the technical characteristics of the assessment when results are released. However, this takes time and can affect the release of NAEP scores, at least for the state NAEP program. Which of the following alternatives do you prefer?

Some technical information should be provided at the time of release, but not as much, so reports can be released sooner (again, about six months). ☐

Full technical documentation is valuable, so the current practice (not releasing some reports until all technical documentation is available) should be maintained. ☐

Comments:

4. INCLUDING A PARENT SURVEY

It may be that some of the information NAEP collects (on parent involvement in schooling, for example) can be provided most accurately by parents. While collecting information directly from parents would provide useful and accurate information, it would raise political concerns and increase burden on schools. Please indicate your opinion about NAEP including a questionnaire which is filled out by parents.

Should definitely have a parent survey ☐

Should probably have a parent survey ☐

Should probably not have a parent survey ☐

Should definitely not have a parent survey ☐

Comments:

5. REPORTING: SCHEDULE FOR RELEASE OF RESULTS

5A. In the past, NAEP results have been released as soon they have been available—at various times during the year, and with different schedules for each subject area. Please indicate whether you think this practice should continue, so results are released when they are available, or whether NAEP should establish a set schedule for release every year.

Current approach ☐

Results should be released at the same time each year ☐

No opinion/Doesn't matter ☐

5B. Should NAEP results be released on the same schedule each year (e.g., , always around 1 February and 1 May), or should NAEP establish a certain time period during the year, such as a week, when all the results from a given year are released? (The latter approach would be like that used by the National Education Goals Panel.) In order to meet the deadlines associated with a regularly scheduled release date, some extra time may have to be built into the schedule, possibly delaying the release of results.

Results should be released on the same schedule each year ☐
Results should all be released during the same week each year ☐
No opinion/Doesn't matter ☐

6. STATE MANDATES FOR PARTICIPATION IN NAEP

For the "national" NAEP, students at grades 4, 8, and 12 are tested in 2-3 different subject areas every two years. For the "state" NAEP, additional samples of students at grades 4 and 8 are tested in selected subject areas. These additional samples permit estimates to be made of student performance in each participating state. Neither NAEP program (national or state) produces results for schools or individual students.

6A. Does your state mandate participation in the state-level NAEP (sampling and reporting to provide state results)?

Don't Know ☐

Yes ☐

No ☐ → **Is your state considering mandating participation in the future?**

Yes ☐ No ☐

6B. Does your state mandate participation in the national level NAEP (sampling and administration to contribute to national-level results)?

Don't Know ☐

Yes ☐

No ☐ → **Is your state considering mandating participation in the future?**

Yes ☐ No ☐

Comments: *(Does your state do anything in particular to encourage participation? We are especially interested in your comments/suggestions about increasing motivation for states, districts, schools, and students to participate in both the "national" and "state" NAEPs.)*

State-level incentives for participation:

District-level incentives for participation:

School-level incentives for participation:

Student-level incentives for participation:

7. SUBJECTS ASSESSED AT THE STATE LEVEL

7A. The following testing schedule has been proposed for the "state" NAEP. For each of the years listed, please indicate whether you think your state would participate in the "state" NAEP program that is offered. (Responses will NOT be seen as commitments by states to participate.)

			Participate?		
Year	Subjects	Grade Levels	Yes	No	Don't Know
1998	Reading	4, 8	Y	N	DK
	Writing	8			
2000	Math	4, 8	Y	N	DK
	Science	4, 8			
2002	Reading	4, 8	Y	N	DK
	Writing	4, 8			
2004	Math	4, 8	Y	N	DK
	Science	4, 8			
2006	Reading	4, 8	Y	N	DK
	Writing	4, 8			
2008	Math	4, 8	Y	N	DK
	Science	4, 8			
2010	Reading	4, 8	Y	N	DK
	Writing	4, 8			

7B. For each of the following subject areas, please indicate whether your state would definitely not participate, probably not participate, probably participate, or definitely participate, if that subject were offered in the "state" NAEP sometime in the future. (Responses will NOT be seen as commitments by states to participate.)

	Definitely Not Participate	Probably Not Participate	Probably Participate	Definitely Participate
Social Sciences and History: Civics, Geography, U.S. and World History, Economics, etc.	DNP	PNP	PP	DP
Arts: Dance, Music, Theater, Visual Arts	DNP	PNP	PP	DP
Foreign Languages: Spanish, French, etc.	DNP	PNP	PP	DP

Comments. *(We are interested in finding out about other subject areas that you think should be included as well as other schedules for testing that you prefer.)*

8. ASSESSING SUBJECT AREAS IN COMBINATIONS OR INDIVIDUALLY

8A. In some subjects, student performance can be measured and reported for the general subject area (such as Social Sciences or Natural Sciences) or for each of the disciplines in the area (such as Civics, Economics, Geography, and U.S. and World History). If results are presented for each discipline, student test time and required resources are greater than if scores are presented at a more general level, and fewer subject areas can be assessed. However, assessing each discipline separately provides greater detail about performance in these specific areas.

If results were reported for more general clusters of subjects, separate scores would still be reported for each discipline or subject in the cluster. However, it might not be possible to report more detailed information, such as sub-area scores in the subject, and it would not be possible to explore the associations of scores with student demographic characteristics or school factors.

For each of the following areas, please indicate whether you feel the subjects in that area should be assessed and reported for each individual subject or skill areas or as a more global combination of the subject areas (clusters).

Social Sciences and History		
Civics	Report as individual subjects	☐
Economics	Report as a cluster	☐
Geography U.S. and World History	No preference	☐
Natural Sciences		
Earth Sciences	Report as individual subjects	☐
Life Sciences	Report as a cluster	☐
Physical Sciences	No preference	☐
Reading/Language Arts		
Reading	Report as individual subjects	☐
Writing	Report as a cluster	☐
	No preference	☐

Comments: (*For example, if you feel subject areas should be organized differently, please indicate this below.)*

8B. Subject clusters can be assessed and reported in an integrated way—in relation to one another. Or, they can be assessed as free-standing areas combined in the same assessment. The former reflects a holistic approach to assessment where tasks are integrated and scored in multiple ways for the subjects involved. The latter approach would combine distinct assessments for each subject and administer them at the same time. Would you prefer that subject clusters be assessed in the integrated approach, or in the "combined" approach?

Integrated Assessment ☐ Combined Assessment ☐ No preference ☐

Comments:

9. DESIRE FOR INFORMATION ON SKILL AREAS WITHIN SUBJECTS

For most subjects assessed by NAEP, it has been customary to provide more detailed results on sub-areas of interest within the overall discipline. For example, in reading, scores are presented for a student's ability to (a) read for literary experience, (b) read to gain information, and (c) read to perform a task. Scores in mathematics are presented in terms of content areas, such as numbers and operations, measurement, and algebra. This detailed information may be very useful to educators or researchers: it can provide data on strengths and weaknesses in various areas to guide curriculum development or teacher training. However, providing it requires resources and may reduce the number of subjects NAEP can assess.

For each of the subject areas and grade levels on the following page, how important is it to you to have subscale (discipline or skill area) scores in addition to overall scores?

(Following are the subscales currently used by NAEP in each subject)

Reading—

Reading for Literary Experience; Reading for Information; Reading to Perform a Task

Writing—

Informative Writing; Narrative Writing; Persuasive Writing

Mathematics—

Number Sense and Operations; Measurement; Geometry; Data Analysis; Algebra and Functions

Science—

Earth Sciences; Life Sciences; Physical Sciences

Geography—

Space and Place; Environment and Society; Spatial Dynamics and Connections

History—

American Democracy; Peoples, Cultures, and Ideas; Economic and Technological Changes; World Role

FOURTH GRADE	Not Important	Somewhat Important	Moderately Important	Very Important
Reading	NI	SI	MI	VI
Writing	NI	SI	MI	VI
Mathematics	NI	SI	MI	VI
Science	NI	SI	MI	VI
Geography	NI	SI	MI	VI
History	NI	SI	MI	VI

EIGHTH GRADE	Not Important	Somewhat Important	Moderately Important	Very Important
Reading	NI	SI	MI	VI
Writing	NI	SI	MI	VI
Mathematics	NI	SI	MI	VI
Science	NI	SI	MI	VI
Geography	NI	SI	MI	VI
History	NI	SI	MI	VI

TWELFTH GRADE	Not Important	Somewhat Important	Moderately Important	Very Important
Reading	NI	SI	MI	VI
Writing	NI	SI	MI	VI
Mathematics	NI	SI	MI	VI
Science	NI	SI	MI	VI
Geography	NI	SI	MI	VI
History	NI	SI	MI	VI

Comments:

10. FREQUENCY OF DATA COLLECTION

For the past several years, NAEP has tested students only every other year. Congress recently authorized the NAEP program to collect data every year. This yearly testing can be implemented in two different ways:

(1) the same amount of testing can occur, but in two years instead of one, or

(2) more subjects could be assessed and/or the same subjects could be assessed more frequently.

If the same amount of testing were done, but on a two year schedule, the burden placed on schools in any single year would be reduced. However, this could increase overall costs and could increase the overall burden placed on schools and districts.

If more subjects were assessed and/or the same subjects were assessed more frequently, much more information could be generated by the NAEP program. However, NAEP costs would be increased substantially and there would probably be more contact and intrusion into local school operations.

Please indicate your opinion on whether (and why) NAEP should collect data every year by checking one of the following options.

Yes -- to reduce the burden each year on individual schools ☐

Yes -- to permit assessments to be conducted in more subject areas or more often ☐

Yes -- to reduce the burden on individual schools and to permit assessments to be conducted in more subject areas or in the same subjects more often ☐

No -- the present schedule should be maintained ☐

Comments:

11. LINKING NAEP TO INTERNATIONAL ASSESSMENTS

Recently, the United States participated in an international assessment of mathematics and science involving over forty countries (the TIMSS Study). The National Center for Education Statistics is trying to link these results to NAEP, so NAEP scores for states can be expressed on the international scale. This could allow comparison of state NAEP scores with the performance of other countries or groups within those countries.

11A. Do you think the government should continue to support this kind of research in the future?

The government should definitely support this kind of research ☐

The government should probably support this kind of research ☐

The government should probably not support this kind of research ☐

The government should definitely not support this kind of research ☐

I have no opinion on this issue ☐

11B. This type of research can be costly, since it is difficult to make a link that is technically sound and credible. Generally, the more that is invested, the more sound the link that can be established. If this research were to continue, at what level do you feel it should it be supported? Should NCES spend just enough to support linkages that might meet minimal standards of validity and reliability (about 0.5% of the total NAEP budget), should NCES spend more to support linkages that might exceed minimal standards (about 1.5% of the budget), or should NCES spend enough to support linkages that would meet very high standards of reliability and validity (about 4-5%)?

Spend enough to meet minimal standards ☐

Spend enough to exceed minimal standards ☐

Spend enough to meet very high standards ☐

Comments:

12. OBTAINING STATE-LEVEL RESULTS

Several options are emerging for generating state-level results from NAEP:

- **Currently, NAEP produces results for individual states by administering the regular "national" assessment to representative samples of students in each state. This approach provides extensive information to states—including results on content area subscales and information on a wide variety of student and instructional background questions—but it is bulky to administer, expensive, requires considerable time to report, and in the future states may have to pay for it.**

- **An alternative to this approach is the so-called "market basket" method, in which representative sets of the assessment exercises would be developed. These would be short modules representing the assessment, which would be offered to states to obtain state-representative results, to calibrate state assessments to the NAEP scale, or to obtain state-level information more often than the NAEP-funded schedule would provide. Standards for administering and using these market baskets would have to be developed, but they probably would be administered in conjunction with the regular national NAEP assessment.**

12A. Which of the following reflects your state's likely interest in using the approaches for obtaining state-level NAEP results?

	Definitely Not Interested	Probably Not Interested	Probably Interested	Definitely Interested
Using the current, full state assessment	DNI	PNI	PI	DI
Using the market basket program	DNI	PNI	PI	DI
Using both	DNI	PNI	PI	DI

Other:_____

Comments:

12B. In addition to the market basket method, other approaches might be developed for providing state-NAEP results. While these alternatives would be easier to administer than the full state NAEP assessment used currently—and probably could be reported much sooner, they would provide less information: nothing on content subscales or background items, for example.

In evaluating alternative approaches for obtaining state-level NAEP results for your state, there are a number of different factors that might be considered. How important are each of the following factors in your state's decisions about assessment approaches?

	Not Important	Somewhat Important	Moderately Important	Very Important
State costs	NI	SI	MI	VI
School burden (number of students tested, staff time)	NI	SI	MI	VI
Psychometric test properties	NI	SI	MI	VI
Ability to compare results with other states	NI	SI	MI	VI
Ability to use results for within state comparisons: gender, race/ ethnicity, public-private	NI	SI	MI	VI
Ability to obtain information on instructional practices and their relationship to student achievement	NI	SI	MI	VI
Ability to obtain student-level results	NI	SI	MI	VI

Comments:

13. STATES PAYING FOR SOME SERVICES

As options are explored for making NAEP more useful to states, a variety of services might be developed. In addition to the labor costs currently associated with the recruitment of schools for participation and the time and ancillary costs required for test administration, how willing would your state be to pay for ALL OR MOST OF THE COSTS associated with the following services? These costs would NOT include test development costs, but would include analysis costs. (This is NOT a commitment; we want to determine how willing states might be to pay for these services.)

13A. A full state-level assessment using the current NAEP approach?

- Would definitely be willing to pay costs ☐
- Would probably be willing to pay costs ☐
- Would probably not be willing to pay costs ☐
- Would definitely not be willing to pay costs ☐

13B. Linking NAEP results with your state's regular assessment program's results, so NAEP results can be estimated from the state program?

- Would definitely be willing to pay costs ☐
- Would probably be willing to pay costs ☐
- Would probably not be willing to pay costs ☐
- Would definitely not be willing to pay costs ☐

13C. Extra "market basket" assessments for various state purposes, over and above those provided by NCES?

- Would definitely be willing to pay costs ☐
- Would probably be willing to pay costs ☐
- Would probably not be willing to pay costs ☐
- Would definitely not be willing to pay costs ☐

Comments:

(OPTIONAL) 14. POTENTIAL AREAS TO MAKE NAEP MORE INNOVATIVE, CREATIVE, EFFECTIVE

There are many areas in which NAEP might make major breakthroughs. This could be in how the subjects are organized, how they are assessed, the working relationship with state and local school systems, or in how results are released or packaged. Please indicate any approaches or ideas you think NAEP should consider in order to make it more useful to you as a customer. Feel free to attach additional sheets as needed.

When you are finished, please return the completed survey (by March 11) to Dr. Roger Levine, A.I.R., P.O. Box 1113, Palo Alto, CA 94302.

APPENDIX B: Summary of Responses to Open-Ended Comments

SUMMARY OF OPEN-ENDED RESPONSES

Question 1: Background Information--Emphasis

Open-ended items reflected the heterogeneity of feelings about the relative importance of various background factors. A few respondents reiterated their feelings that socioeconomic factors were the most important; others were particularly interested in race. On the other hand, a few expressed concern that such student background factors could be used to make excuses and justify mediocrity.

More comments questioned the reliability and practicality of data, especially in the case of instructional practices and topics of current educational relevance. One respondent commented, "We would be concerned about consistent definitions for instructional practices and topics of relevance. It might be important to have the information but it won't be any good unless you can guarantee consistency at some level." While many people felt this data could be extremely helpful, there was concern that it would be difficult to determine effects over time and doubts about the potential precision of links drawn between practices and test results.

In addition, one person complained of repetitive questioning and that the last administration's television questions were redundant. Another requested that the Asian and Pacific Islander ethnic groups be separated. There were also requests for information on homework and other learning reinforcements and an indicator of socioeconomic status more reliable than school lunch program data.

Question 2: Background Questions--Impacts on Release of Results

More people commented on this issue than on any other. While comments were evenly split about the relative importance of background information versus increased timeliness, a large proportion of those respondents who indicated that only "some background information should be collected so reports can be released sooner" still expressed a strong desire for valuable background information. Most notably, more than ten people suggested a "fourth option" of an early release of results with limited background information, with full background information to follow, allowing for the greatly desired timeliness without sacrificing information seen as critical for accurate and responsible analysis.

As in Question 1, opinions were diverse as to the perceived importance of various factors. Valuations of background information ranged from "irrelevant" to "helpful" to "critical to

an accurate analysis and understanding of the final reports." Many respondents were interested primarily in demographic data (race, sex, urbanicity, SES) and felt that other questions are extraneous. Others commented on the "problem of poor interpretation and representation of results" and the need to establish a proper context for NAEP data. It was also mentioned that demographic data is already general knowledge or at least known by schools; this produced the suggestion that certain information be collected only every other administration or every ten years.

In response to the request for specific experiences using these data, a few respondents indicated that data is not used because it is unreliable due to low participation rates or because state rather than NAEP data is used. Most commented that the data has been helpful in setting contexts and comparisons for policy development and in "understanding potential variables that influence student outcomes."

Question 3: Technical Documentation of NAEP vs. Timeliness of Reports

Comments indicated a nearly unanimous concern for quality. Many made a distinction, however, between a majority audience who simply want to be reassured of NAEP's continuing its reputation and tradition of high standards and a minority audience of technical "wonks" who are the only ones who can even understand the bulk of the technical documentation. One member of the majority explained, "Very few people understand the technical side -- fewer use it. We want to be assured that assessment is reliable and valid for our use. We don't need to deal with the technicalities at school and district levels." Speaking for the minority, however, another maintains, "This assessment is widely used to evaluate the state of education in the United States. The validity of the assessment is too important to allow any question which might be answered by the technical documentation (to remain unanswered)."

The overwhelmingly suggested solution was to release limited documentation with a timely report, to be followed later by the full technical documentation. Some respondents drew a distinction such that this later documentation be distributed to a limited audience and for particular purposes. "Research-based reports should not be released without technical backup, but 'report-card' reports can be."

Question 4: Including a Parent Survey

A significant majority of individuals commenting about the implementation of a parent survey were opposed to the idea. While a few of these cited political reasons of

negativity towards federal government involvement in education, most comments regarded more practical considerations of logistics and reliability. Many respondents worried about the "monumental" task of organizing such a survey, and many felt that the financial cost, as well as the necessary energy expenditure, would be prohibitive. Respondents were also concerned with the potential reliability of results, citing problems of parents biased by a desire to "look good" and poor sampling due to self-selection of "available" or involved parents, as well as problems posed in districts where as many as 85 percent of parents are not English speakers. Some felt the burden would even reduce local district participation in NAEP.

There were, however, many respondents who argued for the potential value of a parent survey. They pointed out that parental involvement in education is an increasingly important factor, and a survey would therefore be important for comparing groups and would encourage parents' involvement in their children's educations. Some suggested compromises of conducting a parent survey infrequently, using a limited sample, or as a short survey requesting only the most important information. Still, even some who were in favor of increasing parental involvement felt that such information should be collected at the local level. One pointed out, "(The) purpose of NAEP is to be a 'national report card'--**not** to study parent involvement, however important."

Question 5: Reporting: Schedule for Release of Results

Comments were not explicitly requested for this item, but many respondents chose to share their observations. Though in the objective minority, those who favored a more scheduled approach to the release of results were significantly more vocal. They expressed a desire to know when to expect results so their offices could be better prepared to analyze and disseminate data. One respondent expressed this general concern: "One **huge** problem that I have had is that I have no advance knowledge that NAEP scores are about to be released--therefore, I can't prepare my organization's membership." Some also felt that a schedule would allow NAEP to make a greater impact, and made suggestions for scheduling in order to increase attention from the press.

Many people worried that a set schedule could delay the release of results, maintained that timeliness remains the most important factor to consider and that neither promptness nor quality should be sacrificed. On the other hand, a few suggested that a set schedule might actually speed results by imposing accountability and bolstering the ability to meet a deadline. Others suggested a continuation of the current practice with the addition of a final report that brings together all of the various data previously released.

Question 6: State Mandates for Participation in NAEP

The greatest number of respondents who answered this question indicated that there are no incentives for NAEP participation in their states. A few of these indicated that there was no need for such incentives because participation rates are high in their states anyway. Many respondents noted that participation was encouraged at the state level; "we beg and brow-beat," (we) "call, write, cajole, urge, etc." Some maintained, however, that participation, though requested, should remain voluntary.

The most commonly mentioned explicit incentive was that the state superintendent or commissioner of education writes a letter to school districts requesting participation. Almost as common were financial incentives from states, primarily used to cover the administrative and substitute teacher costs, and sometimes for student level incentives and rewards, such as a pizza party after testing. One state ties overall funding and incentive funds, as well as rewards and the use of distinguished educators as consultants, to school progress in state testing. A few mentioned that districts are provided with testing data and that the validity of NAEP results is sufficient incentive for cooperation.

On the other hand, others noted that districts and schools are not sufficiently compensated for the burden of testing and that NAEP should therefore provide school- and district-level data, or, in one case, financial remuneration. "The 'trial state or state-by-state assessment' needs to provide some method of compensating districts for costs of training teachers…to administer the assessment. NAEP needs either to: a) assume all costs of administration, or b) provide data related to the performance of the participating school to the school."

Question 7A: Subjects Assessed at the State Level

Here again, there were many comments even though they were not explicitly requested. Many small states emphasized that the burden created by testing more than one subject per grade level can affect too many (and possibly every) school in the state, threatening the probability of participation. Some also mentioned their own state assessments as an unknown factor affecting future participation; one added that his state would be more likely to participate in NAEP if they could choose only those subjects that are not already assessed at the state level. A few respondents also mentioned that any cost associated with NAEP testing could hinder participation rates, and others commented that there would be more participation if student-level data were made available. Finally, one respondent noted: "We are still concerned that the same cohort is always tested in the same content areas due to the harmonic created by the 4-year grade span and 4-year

subject rotation." On a more positive note, a few respondents affirmed their intentions to continue participating, one proclaiming, "We value our participation!"

Question 7B: Subjects Assessed at the State Level

The most common response was an expressed preference for focusing on more core curriculum areas. "Our schools participate because core subjects are tested and they view the data as important feedback on their academic programs. If NAEP includes subjects they consider 'non-core academic,' participation will probably drop off." A few respondents disagreed, expressing the belief that any data are helpful, though one still mentioned concerns about achieving participation at the levels required by NAEP, requesting help in providing school-level incentives. Another mentioned that cost would continue to be a factor.

The most common suggestion for other subject areas that should be included was health (or health and safety), which was seen as of particular interest given the current concern with preventing drug and alcohol abuse and changing students' health habits. A couple of respondents also suggested testing in physical education, technology, or computer literacy. Biology, algebra, and vocational education areas were also suggested.

Question 8: Assessing Subject Areas in Combination or Individually

While a few respondents reiterated the overall finding that reading and writing be reported as individual subjects, comments regarding this issue were generally in favor of reporting subjects in clusters. Many noted that such clusters are more in line with their states' current educational approaches. One respondent explained, "our state focus encourages integration within disciplines—the cluster format would enhance this effort." Another expressed a "strong preference for reporting on a cluster level. (This) reflects a more constructivist approach to student achievement and learning." One respondent disagreed: "Poor performance should lead to interpretation or strategies for improvements. The more detailed the analysis of the problem, the easier that becomes." A compromise solution of reporting as individual subjects only every 5-10 years was also suggested.

Only a few respondents commented on Question 8B, most favoring the integrated assessment. One reiterated the trend from Question 8A, saying that the integrated approach supports his/her state's current approach. Another chose the integrated assessment because it would simplify sampling, while someone else felt that it "reflects

what we expect of students in the real world." On the other hand, one respondent chose the integrated assessment but ceded that it would be much tougher to do, and one commented that "(I) don't have enough confidence in the integrated approach in terms of having reliable estimates at the content-specific level."

Question 9: Desire for Information on Skill Areas within Subjects

Slightly more respondents in favor of providing subscale information commented than did those opposed to it. They felt such data would be helpful to educators. One respondent represents the general opinion that "discipline or skill area scores provide data on specific strengths and weaknesses and can be used to support the instructional program." One person noted that the subscales are particularly helpful since they match subareas in the statewide assessment, while another thought that his/her state might align their curriculum to match NAEP's framework. One supporter, however, felt that the detail might only be necessary every 2-3 assessments.

Other respondents were opposed to subscale reporting for a variety of reasons. Some felt that such details are not valuable in state-level results and that NAEP should remain a general benchmark, leaving state and local test programs to execute subscale analysis. A few also felt that the benefits of the information are not worth the cost and burden and may needlessly complicate the reporting process.

Respondents also explained a variety of specific preferences. Several felt that reading is the most important subject to detail, while others focused on writing or math. One suggested the addition of "technical reading and writing." A few respondents felt that younger students are not advanced enough to make subscale data meaningful. Some felt that twelfth-grade data is not useful as results cannot be used "diagnostically" for exiting students; a few suggested testing in the eleventh grade instead.

Question 10: Frequency of Data Collection

This issue evoked many strong opinions, more than two thirds of which were opposed to NAEP collecting data every year. Many of these felt that little or no value would be added by more frequent data. They felt the one-year time interval would be too short to produce new information in any given subject, and a few expressed the common preference (see Question 7B) for limiting assessment to core subjects. Some pointed out that such frequent assessment could lose sight of real trends and become a less substantive tracking of gains and losses in scores. One respondent explained, "Too

frequent testing is not educationally sound; trends take time to develop, while small blips are of little interest." Others expressed little value of NAEP results and felt that without district-or local-level information, the less testing implemented, the better, as it is disruptive and takes time away from instruction. "Data collection should not increase until strong evidence is presented that increased data collection will result in higher student achievement."

Even more people reiterated their concerns regarding the burden placed on schools by NAEP testing, frequently noting that schools feel "over-tested" and that participation rates are therefore threatened by schools' unwillingness. Many, especially those from small states, felt that this schedule change would **not** decrease the burden on schools. One respondent from a small state explained, "Such a large percentage of schools are involved each year, (testing) every year would be an **increased** burden." Others worried about the possibility of the same school being chosen in two consecutive years, given the random sampling technique. Some respondents also mentioned that NAEP is particularly burdensome when added to the workload created by state assessments; one noted that NAEP is used primarily to corroborate state data and that annual testing would be too much for this purpose.

There were, however, a number of respondents who welcomed any additional assessments and increased data. They felt this could "provide a more comprehensive picture of the state of student performance," "help inform instructional practice," and that "more frequent comparisons are required to assess" current changes in curriculum and instruction. There was one request that NAEP address more subject areas (rather than more often), and one respondent felt that annual testing would help smooth the assessment process and increase school participation by lessening the burden.

Question 11: Linking NAEP to International Assessments

Even though most respondents were in favor of supporting this kind of research, most of those offering comments had concerns or reservations about this kind of research. Some were concerned about implementation of these linkages, interpretation of results, and other research issues. Concerns were raised about "biased and political" sampling; about the feasibility of conducting this linking "in a credible manner," and about the lack of comparable context variables. One respondent questioned the value of this kind of international research by writing, "Not sure how much it matters how we compare to Finland and such."

Other respondents felt compelled to elaborate on their support of this kind of research. Two respondents commented on the global economy. "The U.S. is in competition with other countries, so it's important to know how U.S. students are performing in relationship to students from other countries. In fact, restricting statistics to comparisons among states is misleading. A state that is doing well in comparison with other states may reduce its attention to education, when in reality it is still behind other developed nations."

Question 12A: Obtaining State-Level Results

The "market basket" concept evoked interest from several respondents; however, there was concern that more information was needed to permit a better evaluation. Respondents also used this item as an opportunity to express concerns about the ancillary costs associated with NAEP and the importance of this factor in their decisions.

Very few respondents saw a need to comment on using the current assessment in this item.

Question 12B: Obtaining State-Level Results (Factors in Evaluating Alternatives)

Most of the respondents commented on the "Ability to obtain student-level results" factor. Four respondents commented favorably about being able to obtain student- (or school-) level results. These comments were countered by other respondents who raised the points that, "This question is out of the parameters of NAEP" or they definitely were not interested in student-level results. There was clear interest in district-level results, and how this would make NAEP more attractive to school districts.

Other factors were elaborated on by one or two respondents. It was interesting to note one respondent's comments about psychometric test properties: "From a policy standpoint, we assume the test writers have taken care with psychometric test properties, so we are concerned more with other issues." The relative absence of comments about costs and school burden undoubtedly reflects the fact that respondents for whom these factors were important were provided many opportunities to make such comment in response to previous items.

Question 13: States Paying for Some Services

Several respondents voiced their frustration at being unable to answer this item -- "Need details," "Depends on how much," "Depends on total money," "Hard to tell. So many factors are involved" were responses typical of these kinds of respondents.

The majority of comments were related to costs. Some respondents explained their tight budget situations: "Our state is in **very** difficult financial times. Education fights for every penny... It's painful." "We cannot even get state funds for what we as a state should be doing." Another volunteered that "Cost is probably why we don't participate in the first place." One respondent who indicated his state would be unwilling to pay costs explained that, "Our response really reflects being 'able' to pay costs. It does not reflect a lack of valuing NAEP results or wishing to calibrate with NAEP."

14. Potential Areas to Make NAEP More Innovative, Creative, Effective

This optional item, as expected, elicited a wide range of comments. Some respondents prepared and attached well thought-out comments. Another respondent attached his district's frameworks for teaching and learning in its public schools.

Several respondents commented on the value of linking NAEP with their state's assessment. Other commonly mentioned areas included a desire for a more rapid dissemination of results, the absence of rewards for schools for participating, and a desire to use this information to improve the quality of instruction.

We found it rewarding that one respondent commented, "In order to provide more useful information to states and localities, there could be a forum or methodology for more interaction between NAEP and localities, such as this survey.... This sort of information gathering would be helpful."

APPENDIX C: Focus Group Protocol (Media Group)

NAEP Media Focus Group Guide – 29 April 97 draft

For the focus groups, do the following:

-- ask participants to arrive promptly at the scheduled time

-- have **two staff members run the group**, taking turns as appropriate, with one leading the group and the other recording main ideas and seeking clarification as needed

-- make reminder calls the preceding day

Have ready:

-- a table with **coffee, juice,** and **food**

-- **handouts** of schedule, sub-scales, and subject clusters

-- **tape recorder**, to tape the session. Don't forget to get people's agreement before using. Explain that we want to tape the session to allow accurate representation of their feelings. Assure the respondent that everything is confidential and that the tape will be erased after our notes are prepared.

-- large **name cards** that can be placed in front of each participant, with first name in BIG letters; last name in smaller ones

-- a watch or clock, so you can monitor progress against the anticipated schedule.

Format

Introductions. *(10 minutes)* Introduce yourself. Have writers introduce themselves, identify the publications they write for, and briefly describe the type and extent of education writing they do. Establish ground rules (e.g., no "right" answers; everyone gets a chance to talk). Remind people that their responses will be taped (for record keeping purposes only) and that everything is strictly confidential.

NAEP background. *(5 minutes)* Introduce the task, emphasizing these points (written as a script but to be done in our own words.):

-- Congress mandated the National Assessment of Educational Progress in 1969. Since then, it has regularly monitored performance of 4th, 8th, and 12th grade students in reading, writing, math, science, and history/geography.

-- NAEP is also known as "The Nation's Report Card". It is not the new student assessment program Clinton mentioned in his inaugural address. No one knows what this new test will be like.

-- NAEP is the only long-term, longitudinal national study of student achievement.

-- Until 1988, NAEP could only report results on a national level. Congress then authorized "trial state assessments" to allow the presentation of results at the state level.

-- NAEP scores are reported on a 0 - 500 point scale. Results are also reported as the percentage of 4th, 8th, and 12th graders performing above different proficiency levels (advanced, proficient, basic) or below basic.

-- NAEP results are often covered by the press, and one of the issues that NCES is interested in is how the media feels NAEP reporting can best meet their needs.

Objective. *(3 minutes)* NAEP is being redesigned. We want to know how writers feel about some of the alternatives under consideration and how you feel NAEP should be changed. Assure participants that by participating in the focus group, they are not supporting NAEP, but instead helping NAEP better design its reports for greater accessibility and usefulness.

The following is the currently anticipated national and state-NAEP testing schedule. *Show handout with National and State NAEP schedule.*

Desire for information on skill areas within subjects. *(10 minutes)* NAEP is now designed to provide detailed results on sub-areas within each of these subjects. *Hand out sheet with list of subscales within each subject area.* As you know, doing this costs money and may cut down on the number of subjects that can be measured or how often they are measured. How do you feel about this — is this information useful for reporting? Is it worth the costs?

Are there any grade levels for which you feel this information is more important or useful to report?

(Not a survey question): **Are there other subjects for which you feel national test scores should be produced?**

Assessing subject areas in combination or individually *(5 minutes)* Let's begin with a discussion of some of the subjects that are included in NAEP and how you'd like scores in these subjects reported. *(Distribute handout with subject areas and individual subjects — as listed in item)* NAEP can be designed to produce a score for "Social Sciences and History" — or it can be designed to provide detailed scores for Civics, Economics, Geography, and U.S. and World History. There are costs and benefits associated with either approach.

Reporting scores for each subject area provides greater detail and allows investigation of student and school characteristics that might be associated with each of these subject areas. But, it takes up NAEP resources that might be used for other purposes. Conversely, reporting combined scores, which would still report subject scores, would not permit breaking up subject scores into finer scores or investigating school and student characteristics which might be associated with subject area scores.

Which would be more useful for you — a combined Social Sciences and History score, or having NAEP produce subject area scores which could be studied in greater detail? Why?

How about Natural Sciences: Which would be more useful for you — a combined score or individual scores for Earth Sciences, Life Sciences, and Physical Sciences? Why?

How about Language Arts. Which would be more useful for you — a Language Arts score or Reading and Writing scores, which could be studied in greater detail? Why?

Background information - emphasis. *(10 minutes)* NAEP collects background information -- about school characteristics (like size, location, and type), instructional practices and techniques used at the school (like content coverage and teaching technique), and student background characteristics (like race/ethnicity, gender, parental involvement, TV viewing) -- to explore relationships between these characteristics and achievement. NAEP could also ask about topics of current educational concern, such as portfolio assessment, year-round schools, and charter schools). What sorts of things would you, as a writer, be most interested in reporting about? We're talking about investigating the relationships -- or lack of relationships -- between these factors and achievement. What is the most important information for NAEP to provide?

(If not mentioned, ask about each of the specific areas below):

School characteristics -- size, location, grade ranges served -- are things that might be associated with achievement. How informative is this information for you as a writer?

How about *instructional practices* -- or other areas of current educational relevance. How interested are you in having this information available? Why?

How about student and home background characteristics -- race, gender, parental

involvement -- how interested are you in having this information available? Why? How could you use this information?

Parent survey. *(5 minutes)* Parents are probably the best source of information about the things they do that might be associated with student achievement. One way to get this information is to survey parents directly. However, this would raise political concerns. How do you feel about the idea of adding a parent survey to NAEP? Why?

Background questions: Impacts on release of results. *(7 minutes)* There are costs associated with all decisions. Overall, is it worth collecting background information if it delays the release of NAEP scores?

What is a reasonable delay? For example, if 1996 test scores could not be released until 1998, this clearly would be too long.

Impacts of results (general). *(5 minutes)* What impacts does the release of NAEP results have on your community? For example, what happens when the government releases information that shows how much the typical American student or the typical 8th grader in your state knows about Geography. What effects does this have on your readers?

Release of results. *(5 minutes)* Presently, NAEP releases reports as they become

available. What if NAEP results were always released at the same time each year. Would this make NAEP more useful for you?
IF SO, What would be the best time of year for the release of results.

Technical documentation: Impact on Timeliness of reports. *(10 minutes)* When NAEP reports are released, they include a lot of technical documentation. If NAEP didn't have to wait until all of this technical information was calculated, reports could be released about six months earlier. What are the advantages and disadvantages of delaying release six months, until there is complete psychometric documentation?

Do you feel it makes sense to delay release of results for six months for this reason? Why (not)?

Linking NAEP to international assessments. *(10 minutes)* Recently, results from the Third International Math and Science Study, or TIMSS, were released to the public. It may be possible to link NAEP Math and Science scores to these TIMSS scores. This would allow comparisons of state NAEP scores with the performance of other countries or groups within those countries. This has been done in the past (pass out example)

Should the government be supporting research to do this? Why or why not?

STATE-NAEP: Would you like your state to or do you think states should link or calibrate its assessment program to NAEP, so your community's results on that program could be expressed on the NAEP scale? Why or why not?

Ending. What kinds of comparative information about student achievement are most useful to writers?

United States
Department of Education
Washington, DC 20208-5652

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G-17

Fourth Class Special
Special Handling