

148936



United States  
General Accounting Office  
Washington, D.C. 20548

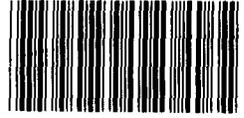
Program Evaluation and  
Methodology Division

B-252853

April 12, 1993

The Honorable William D. Ford  
Chairman, Committee on Education and Labor  
House of Representatives

The Honorable William F. Goodling  
Ranking Minority Member  
Committee on Education and Labor  
House of Representatives



148936

The Honorable Dale E. Kildee  
Chairman, Subcommittee on Elementary, Secondary,  
and Vocational Education  
Committee on Education and Labor  
House of Representatives

The Congress is considering dramatic steps toward the reform of elementary and secondary education that would encourage new standards for teaching and learning, new assessments of student learning linked to those standards, and new policy and oversight bodies and procedures to approve standards and assessments and to report educational results. In response to your earlier request, we are completing a number of reports on national standards and testing. This letter responds to your recent request that we review the proposed policy and oversight arrangements.

Below, we review the background to the request, including the specific questions you asked and our approach to answering them. We then summarize our results in brief, present the framework that guided our work, and discuss the proposed arrangements and the implications we see for congressional action. We conducted our work in February and March 1993 in accordance with generally accepted government auditing standards.

BACKGROUND

In January 1992, the National Council on Education Standards and Testing (NCEST) reported to the Congress and the American people that the absence of standards for educational

05639/148936

achievement and of assessments that measure student progress in terms of such standards has hampered American education. NCEST recommended that a system of standards and assessments be created under the oversight of the National Education Goals Panel (Goals Panel).<sup>1</sup> The system would include (1) voluntary national content standards that would specify the challenging material that students should know in order to meet national educational goals; (2) voluntary national school delivery standards that would identify the school resources and conditions necessary to give students a fair opportunity to learn the material specified in the content standards; and (3) a system of assessments--including large-scale national assessments and assessments at the state, district, and classroom levels--to measure students' attainment of the new standards.

The Congress responded to NCEST's recommendations by drafting legislation to lay the foundation for a system of standards and assessments.<sup>2</sup> Under this proposed legislation, an expanded Goals Panel would certify the voluntary national content standards and school delivery standards. A new National Education Standards and Assessments Council (NESAC) would review content standards proposed by subject-area groups and would report to the Goals Panel. A separate group would develop school delivery standards and propose them to the Goals Panel. The National Academy of Sciences (NAS) would develop criteria for the review of assessments. (The composition and duties of these and other bodies are summarized in enclosure I.) The proposed legislation was considered in both Houses in the last session but did not reach final passage.<sup>3</sup>

You have indicated interest in moving expeditiously with legislation, but also in ensuring that longer term issues are anticipated as the first steps of implementation of the standards and assessment system are planned. You requested that we draw on

---

<sup>1</sup>NCEST, Raising Standards for American Education (Washington, D.C.: January 1992). The Goals Panel was established to monitor progress toward meeting the national education goals selected at the meeting of the President with the nation's governors in Charlottesville, Va., in September 1989. Its members include governors, Members of Congress, and representatives of the administration.

<sup>2</sup>Ideas from both House and Senate were merged in part A of title VIII of the proposed Neighborhood Schools Improvement Act. U.S. Congress, Conference Report to Accompany S.2, House Report 102-916 (September 25, 1992).

<sup>3</sup>A similar bill, H.R. 92, was introduced into the 103d Congress on January 5, 1993, and was referred to the Committee on Education and Labor.

our body of work on testing to consider several questions concerning the arrangements proposed in S.2 and their implications.

- What functions will likely need to be performed, in what sequence, to certify proposed voluntary national education standards and to create a system of assessments?
- Are the National Education Goals Panel and the new National Education Standards and Assessments Council, as currently proposed, appropriately constituted to perform these functions?
- How will the existence of national bodies charged with certifying national standards and assessments affect the National Assessment of Educational Progress (NAEP) and the responsibilities currently assigned by statute to its governing board? What steps can be taken to avoid overlap and conflict of responsibilities among the National Assessment Governing Board (NAGB), the National Center for Education Statistics (NCES), the Goals Panel, and NESAC?

We have already provided testimony before the Subcommittee on Elementary, Secondary, and Vocational Education this year summarizing our overall observations on general issues of standards and testing drawn from our work.<sup>4</sup> To respond to your request, we did not gather new data but reviewed the evidence, analyses, and conclusions from our three studies concerned with testing. These are:

- a survey of current U.S. testing practices and of state testing officials' opinions about assessment issues and national testing;<sup>5</sup>
- a study of the Canadian experiences with standards and assessments that are similar in many ways to the ideas being considered here;<sup>6</sup> and

---

<sup>4</sup>Student Achievement Standards and Testing (GAO/T-PEMD-93-1, February 18, 1993).

<sup>5</sup>Student Testing: Current Extent and Expenditures, With Cost Estimates for a National Examination (GAO/PEMD-93-8, January 13, 1993).

<sup>6</sup>This report is forthcoming.

-- an evaluation of NAGB's experience in developing national achievement standards for judging students' performance on NAEP.<sup>7</sup>

These studies examined issues of how standards are and should be set, how tests can reflect standards, and how oversight groups in this country and Canada have so far handled such matters. The latter two studies have given particular attention to how oversight bodies have used expert knowledge in planning for assessments and in evaluating whether they provide fair and accurate measures of achievement of standards. Such concerns about tasks and capacity are at the heart of the new questions the Committee posed.

#### RESULTS IN BRIEF

We identified four major functions that will need to be performed in sequence to achieve a system of assessments that measure students' attainment of national standards: (1) review and certification of the content standards; (2) design of the assessment system, with special attention to the role of national testing; (3) review of particular assessments within the system; and (4) evaluation of the new system's effects. The proposed legislation covers primarily the first of these functions; that is, content standards review. It gives some attention to preparing for assessment review (the third function). However, our analysis suggests that more should be done at this first stage to anticipate issues of measurement and of system design that will arise as standards are translated into assessments; to examine possible mechanisms for reviewing assessments; and to lay the foundation for evaluation.

The Goals Panel envisioned in S.2 is a high-level policy body that is appropriately constituted to certify standards and assessments, but it will rely on NESAC to ensure that all relevant concerns about proposed content standards have been identified. With only one or two measurement experts among its members, NESAC is not appropriately constituted to identify measurement concerns. Review of assessments also requires specialized knowledge; no specific mechanism for performing this function has yet been proposed.

NAEP provides the only nationally representative measure of student achievement, and policy decisions on NAEP will be critical to the design of the assessment system. The major governance issue is whether these decisions should be made at the level of the Goals Panel, rather than by NAGB. Additional issues

---

<sup>7</sup>We reported interim findings in National Assessment Technical Quality (GAO/PEMD-92-22R, March 11, 1992). The full report of our work is forthcoming.

will arise with respect to NAGB's responsibilities to determine test content, to identify achievement goals, and to set policy for linkages between NAEP and other tests. Taken together, these issues suggest a need to redesign governance arrangements for NAEP.

FUNCTIONS AND TASKS TO BE PERFORMED: OUR FRAMEWORK

As the basis for our analysis in this report, we created a framework of functions and tasks needed to create a workable system of assessments based on challenging content standards. We began by considering what is likely to happen as standards are certified and as states voluntarily adopt and implement those standards. From our overall body of information on past and projected testing in this country, as well as related experience in Canada, we derive three key observations:

- Challenging content standards are likely to include skills and knowledge that many schools cannot now deliver and that cannot readily be measured by current large-scale assessment techniques.
- The content standards will be voluntary; thus, different states will likely adopt different "packages" of skills and knowledge from the national standards. (The more the national standards identify options as well as core knowledge and the more they exceed current school delivery capabilities, the more diverse the state choices are likely to be.)
- Each "package" of standards will likely generate several different assessments, each of which measures aspects of the content standards in accordance with a particular purpose and method of measurement. Making a "system" of these diverse assessments will be no easy task, and one that will require considerable planning.

Reviewing the proposed legislation in light of these observations, we saw that content standards review and assessment review were not the only major functions to be performed: attention to assessment system design would also be needed. We added evaluation as a fourth major function that should always be included when any major new policy is introduced. We identified particular tasks needed to perform or to prepare for each of the four functions and created the framework shown in table 1. The sections that follow discuss each major function.

Table 1: Framework of Functions and Tasks<sup>a</sup>

Major function	Tasks for this function	Tasks for future functions
1. Select content standards	Review content standards (NESAC) Certify content standards (Goals Panel)	Identify assessment feasibility issues Review delivery implications by subject (school delivery standards group) and cumulatively Plan for evaluation and collect baseline data
2. Design the assessment system	Set policy for NAEP and any other national assessment, for linkages to other tests	Develop and test new measurement techniques Design assessments and linkages between assessments Collect baseline data and other evaluation data
3. Review assessments	Review assessments <sup>b</sup> Certify assessments <sup>c</sup> and report progress toward meeting national goals	Collect baseline data and other evaluation data
4. Evaluate the effects of standards	Analyze baseline data and other evaluation data to determine the effects of the standards and assessments	

<sup>a</sup>Responsibilities proposed in S.2 are shown in parentheses. Absence of parentheses indicates that no entity has been assigned this task.

<sup>b</sup>Possible NESAC responsibility.

<sup>c</sup>Possible Goals Panel responsibility.

## CONTENT STANDARDS REVIEW: TASKS AND ARRANGEMENTS

S.2 provides for the core task of content standards review by instructing NESAC to determine whether proposed standards: (1) are sufficiently general to be adopted by any state, (2) reflect the best evidence available regarding the knowledge and skills that students should acquire, and (3) are sufficiently challenging to ensure that American students receive instruction at world-class levels.

Our framework identifies additional tasks that should be performed at this stage to prepare for future functions. Having NESAC perform these tasks as part of a coordinated review will help ensure that the Goals Panel's standards certification decisions are fully informed.

### Assessment Feasibility Review

Review of the match between proposed content standards and what we know how to measure in education should be incorporated into the content standards review. Assessment feasibility is not necessarily a criterion for approval of content standards; credible and challenging standards can be recommended even if no satisfactory method for measuring whether students have achieved them currently exists. However, attention to measurement issues at this early stage can be helpful in preventing later difficulties. Such attention can (1) alert educators, the public, and assessment developers that parts of standards include knowledge or skills for which measurement should be attempted with care and with limited stakes for students or teachers until measurement quality is established; (2) provide a guide to needed research and development; and (3) identify issues that are likely to require special scrutiny as assessments are reviewed.<sup>8</sup>

### Delivery Feasibility Review

Second, attention to the practical feasibility of the content standards--that is, identification of those portions of the standards that schools currently lack the capacity to deliver--should be included in the review in anticipation of adoption and assessment issues. Judging from the degree of concern over local control expressed by state testing directors in response to our recent survey, we know that feasibility will be a factor in state decisions about whether to adopt the standards wholly or in part. Feasibility and opportunity-to-learn considerations will also

---

<sup>8</sup>Attention to measurement at this early stage can also help to identify portions of the proposed standards that are unclear or ambiguous. If experts cannot agree on what they should try to measure to see if students have mastered a standard, the standard may need refinement.

help determine which aspects of the standards are incorporated into assessments in states and at the national level.<sup>9</sup>

We could not clearly determine from S.2 that delivery feasibility would be part of NESAC's content standard review responsibilities.<sup>10</sup> (The requirement that content standards be "sufficiently general to be adopted by any state" could imply attention to practical considerations, but it could be also be interpreted to refer only to avoiding inappropriate narrowness of content.) To encourage analysis of feasibility as early as possible while subject-area groups are developing standards, the Goals Panel and NESAC might request those who submit standards for review to include feasibility information along with other supporting data.

Further, as standards are certified, the Goals Panel will need to assess the cumulative impact that the growing set of standards may suggest for the nation's schools. Cumulative impact should not be a factor in the certification of any one set of standards. However, attention to cumulative resource requirements (net of overlap between related subjects such as history, geography, and social studies) is critical in terms of state adoption and implementation. The Goals Panel and the public need to be informed whether, taken together, national standards can or cannot be accommodated within the time and other resources generally available for education.<sup>11</sup>

#### Implications for Capabilities and Coordination

Our expanded list of functions has two main implications: one concerns the background and qualifications of reviewers; the other, the coordination needed between the review of content standards and the development of school delivery standards.

---

<sup>9</sup>Canadian provinces ensure consideration of school delivery issues as standards are developed by relating standards to the curriculum and by major involvement of classroom teachers. Standards developed by national groups of experts may vary in the degree to which they take practical considerations into account.

<sup>10</sup>The school delivery standards consortium has a related responsibility, as discussed below.

<sup>11</sup>If standards include optional as well as core skills, a modest degree of excess of cumulative demand over cumulative resources is not worrisome. Each state will choose some optional skills and not others, in accordance with its own needs and priorities. An excess of core demand could be a problem, however.

## Reviewer Capabilities

In making standards certification decisions, the expanded Goals Panel (which will be comprised of governors, state legislators, Members of Congress, and representatives of the executive branch) will rely on NESAC to ensure that all relevant technical and nontechnical concerns about proposed content standards have been raised and have been adequately considered. As proposed, NESAC is to include 15 members: 5 educators, 5 individuals from the general public (such as public officials and representatives of businesses, foundations, and advocacy communities), and 5 "individuals with expertise in educational assessment, content standards, and curriculum design" (see enclosure I). This composition appears appropriate for the performance of the content standard review tasks specified in S.2, which are not of a technical nature.

However, the proposed composition of NESAC will not be adequate for the assessment-related activities that we believe are necessary at the content standard review stage. Identifying the difficulties that are likely to arise in assessing student attainment of a content standard requires specialized knowledge of measurement issues whose significance often is not apparent to nonexperts. With five positions allotted to cover assessment plus two other important areas of expertise, NESAC might well include only one or two members who have this specialized knowledge.

Our analysis of NAGB's operations is relevant because, like NAGB, NESAC has both policy and technical responsibilities. We think that, like NAGB, NESAC needs more members with technical expertise to ensure that measurement issues are raised and that their significance is understood.<sup>12</sup> Without such representation, NESAC risks overlooking or setting aside technical issues that can cause problems later on.<sup>13</sup> Making knowledge of assessment issues a priority for more of the NESAC

---

<sup>12</sup>Like the proposed NESAC, NAGB represents educators, officials, and members of the general public. Two of the 23 NAGB positions are reserved for experts in testing and measurement, but during the period we studied, only one of these individuals was an expert in issues of assessment design.

<sup>13</sup>NESAC's capabilities for reviewing assessment feasibility could be extended by (1) forming an assessment task force of NESAC members and other experts, (2) providing for technical staff or consultants to NESAC, or (3) arranging for assistance from NAS, which under S.2 would also be developing criteria for assessment review. We favor option (1), on the grounds that involvement of NESAC members will help ensure effective communication between experts and nonexperts.

membership will help to build confidence that technical weaknesses will be foreseen and averted.

### Coordination With School Delivery Standards

We argued earlier that practical feasibility is a key added dimension for review of content standards. The plans in S.2 call for feasibility issues to be examined, but they assign this task to a separate body, a consortium formed to develop school delivery standards. The consortium is to identify the resources, practices, and conditions that are needed to give all students a fair opportunity to achieve the knowledge and skills specified in the content standards and to propose these as delivery standards sufficiently generic to be adopted by any state. The Goals Panel will consult with NESAC once these standards have been proposed for certification, but we were unable to determine whether or how the development of delivery standards would be coordinated with the review of content standards prior to this final point.

We see the two groups' work as interdependent, and we are concerned that lack of coordination could result in lack of correspondence between the two sets of standards, particularly if delivery feasibility is not considered during content standard review. For example, the school delivery consortium might determine that an aspect of the standards for geography requires a specialized resource that is not widely available and that no delivery standard corresponding to this content standard should be established. This would leave us with nationally approved content that schools are not expected to teach.

Solution to this problem could take several forms. NESAC review could be designed to ensure that major delivery feasibility issues are identified prior to Goals Panel certification of content standards, and delivery standards developed thereafter. Alternatively, the Goals Panel and NESAC could ask the delivery standards group for a feasibility review of each proposed content standard. In either case, it seems clear that a coordinated approach to review and reporting could better inform the Goals Panel and the public of the overall strengths and weaknesses of proposed content standards. A final option would be to develop both content standards and delivery standards without regard to practical feasibility and to leave it to each state to adopt delivery standards consistent with the content standards it has adopted for itself.

### ASSESSMENT SYSTEM POLICY AND DESIGN

We identified a second broad function not addressed in the activities authorized in the proposed legislation, but precisely the sort of longer term matter mentioned in your letter requesting this review. That is, a national system of assessments needs deliberate, proactive design; it will not

emerge from the collage of state-adopted segments of national content standards and clusters of states developing diverse measures. Since the future role of the National Assessment of Educational Progress is one of the most central questions of policy in developing our national system of assessments, we defer our discussion of this function until we reach the Committee's second question about coordination of NAEP governance with the new entities.

### REVIEW OF ASSESSMENTS

Our third major function, which was recommended by NCEST and addressed in part in S.2, is the review of assessments to determine whether they measure student achievement in terms of the new standards and are of high technical quality. As we noted in our initial key observations, challenging standards are likely to involve skills and knowledge that are beyond current measurement capability and will likely result in a diverse array of assessments. We examined whether arrangements currently proposed for reviewing content standards would be sufficient to perform the technically demanding assessment reviews implied by these observations.

#### The Assessment Review Task

Expert analysis is the usual basis for evaluating alignment of assessments to content standards and for evaluating their technical quality. Alignment of test content is typically reviewed by subject-area specialists; that is, people who can read each test item, judge the skills and knowledge needed to answer the item correctly, and relate the skills examined by the test to the content the test purports to cover. For example, experts in mathematics have reviewed the NAEP mathematics test to assess how it reflects the curriculum standards adopted by the National Council of Teachers of Mathematics. Alignment can be assessed before administering a test.

The technical quality of a test may be judged in part by examining test items and plans for test administration and scoring. However, evaluation typically rests on evidence derived from analysis of testing practices and test results. For example, fairness may be judged by examining whether the test was given and scored under uniform conditions, and statistical checks are applied to see whether responses to test items are differentially affected by students' gender, race, or ethnic membership. Determinations of test validity also typically rely on inspection of test results in comparison to other indicators of student performance and require examination of whether test data are adequate for the purposes for which they are to be used. Evaluation of claims of linkage (comparability of results) between one test and another requires an understanding of test

construction, test administration, and any statistical transformations involved.

### Arrangements for Review

Based on our past observations of panels in the United States and Canada dealing with technical testing topics, we concluded that NESAC is not appropriately constituted to conduct assessment reviews. To begin with, few members of NESAC will likely be measurement specialists or experts in the subject covered by a particular test. Secondly, the number of assessments needing review and certification--national, state, regional, commercial, and perhaps district-level assessments in many subject areas--could be quite large. NESAC is not designed to deal with a constant stream of applications for technical review.

We can envision several possible assessment review structures including: (1) a single organization performing technical reviews, perhaps selected by competition for a contract from the Department of Education to support the work; (2) several such centrally selected entities; or (3) a fully decentralized model in which each testing entity gets its tests reviewed by an independent, qualified organization with psychometric and other necessary expertise. We think the NAS could be asked to review the pros and cons of various structures. The need to decide upon a review structure does not seem particularly urgent since it will be some time before any content standards are approved and since it will take additional time to develop and refine assessments aligned to those standards.

### Criteria for Reviewing Linkages Between Assessments

The proposed legislation provides for NAS to propose criteria for evaluating assessments that claim to be aligned to national content standards. The NAS review is to address criteria concerning: (1) whether the subject matter of the test adequately reflects the aspects of the content standards it claims to incorporate, and (2) whether the assessment is of high technical quality--that is, whether it produces reliable measures of student performance and is fair--and is valid for its intended purposes. From our observation that difficult and complex standards will result in states' adopting different packages, we note that assessments of these different packages will inevitably cover different material and may use different methods. This likelihood, in turn, raises difficult questions of linking the results of the various assessments to each other so that, for example, one could say that a score of X on a state test was

comparable to a score of Y on NAEP.<sup>14</sup> Criteria thus should also be developed for evaluating whether specific plans for linkage are sound and whether interpretation of results from one test in terms of another is valid.

### EVALUATION

The proposed legislation does not address our fourth major function, which is evaluation of the effects of the voluntary national content standards and system of assessments. S.2 gives the Goals Panel a general responsibility to propose baselines and benchmarks against which progress toward achieving the national education goals can be evaluated, but makes no specific provision for examining the effects engendered by the adoption of national content standards.<sup>15</sup> If no provision is made for evaluation, valuable lessons about the intended and unintended consequences of this new system--and how and why they occurred--will be lost. At present, we are making a national assumption that standards improve education and that assessment plays a major role in such improvement. Thus, we must examine whether or not--and to what degree and under what circumstances--that proves to be the case.

Plans for an evaluation, if there is to be one, should be laid early and should provide for the collection of baseline measures that can be used to track changes in educational practice and in student performance on a broad spectrum of skills, not simply those highlighted in the standards. We return to this point in our discussion of national testing, below.

### SUMMARY OF OUR CONCLUSIONS REGARDING FUNCTIONS AND CAPABILITIES

Our conclusions with respect to the proposed legislation are as follows:

1. Content review. Two new tasks--assessment feasibility review and delivery feasibility review--should be performed along with the content standard review in order to prepare for assessment design. In reporting to the Goals Panel and to the public, the review body should identify those portions of the

---

<sup>14</sup>Robert J. Mislevy, Linking Educational Assessments: Concepts, Issues, Methods and Prospects (Princeton, N.J.: Educational Testing Service, December 1992). Reviewing the state of the art, the author argues that linkage must be considered from the outset as different tests are designed. "There are no neat technical tricks by which the results of just any old assessments can be compared" (p. i).

<sup>15</sup>The proposed language does provide for NAS to evaluate the effectiveness of the voluntary national school delivery standards, but not the content standards.

standards that raise assessment issues, either because schools do not generally provide students with the opportunity to learn the material or because no satisfactory method of assessment is yet available.

NESAC is not adequately constituted to undertake these tasks. Technical representation on NESAC should be expanded, and decisions should be made about how to coordinate NESAC's work with that of the consortium that is formulating school delivery standards.

2. Assessment system design. Proposed legislation does not address this function. Proactive design of the assessment system, including decisions about the role of a national test such as NAEP, will be needed.
3. Review of assessments. Assessment review is primarily a technical function for which NESAC is not suited. NAS could be asked to develop options for securing the necessary technical evaluations and criteria for reviewing linkages between assessments.
4. Evaluation. No provision has yet been made for evaluation of the effects of the new system of standards and assessments. Plans for evaluation should be laid early.

The National Assessment of Educational Progress will likely be a key component of the assessment system and also of an evaluation plan. Thus, decisions about NAEP will be closely intertwined with decisions reached by the Goals Panel. We now examine the implications of this new situation.

#### GOVERNANCE ISSUES: NAEP AND THE NEW ARRANGEMENTS

In response to the Committee's final question, we now turn to the NAEP and to the issues of governance and coordination that will arise as the system of standards and assessments, with NAEP as a central source of national data on student performance, is created. We believe the chief issue is whether policy decisions concerning NAEP's purposes, subject coverage, and connection to other assessments and to evaluation should remain with NAGB or whether these decisions should now be made at a higher level, as part of system policy and design (a major function on which S.2 was silent). Subsidiary issues will arise with respect to specific responsibilities now assigned to NAGB or NCES. Following a brief description of NAEP and its current governance, we will discuss these issues in turn.

#### NAEP and Its Governance

NAEP is a federally supported student testing program that is now the only source of nationally representative data on student

achievement in basic school subjects. By statute, NAEP tests samples of students in grades 4, 8, and 12 in reading and mathematics (every 2 years), writing and science (every 4 years), and in history and geography or other subjects (at 6-year intervals). Traditionally, NAEP tests have covered a broad range of commonly taught material (from very easy to reasonably difficult for the grade); NAEP now seeks to reflect emerging practices to some degree as well. In accordance with NAEP's statutory purposes, the tests have been designed to show the range of achievement (from least to most proficient performance), to measure average performance accurately, and to monitor changes in performance over time. Since 1990, results have been reported in terms of performance levels or goals established by NAGB. For 1990 and 1992, state-level NAEP assessments were authorized on a trial basis.

Policy guidance for NAEP is provided by the National Assessment Governing Board, an independent body established by statute in 1988 to represent NAEP's varied constituencies. Two of the 23 NAGB positions are reserved for individuals with expertise in testing and measurement. NAGB's responsibilities pertinent to the standards issue include (1) selecting the subjects to be assessed, beyond those mandated by statute; (2) approving the framework of content to be covered in each assessment, as recommended by "consensus groups" of educators and citizens, (3) identifying appropriate achievement goals, and (4) developing standards and procedures for interstate, regional, and national comparisons. (NAGB's composition and functions are summarized in enclosure I.)

NAGB's guidance is carried out by the National Center for Education Statistics, which is the statistics branch of the Department of Education. NCES oversees the technical contractor that administers the assessment and prepares the results for reporting. NCES is responsible for ensuring that NAEP presents achievement fairly and accurately and that published results meet statistical quality criteria. The Commissioner of Education Statistics, who heads NCES, is responsible for conducting evaluations of NAEP. (The overall governance structure of NAEP is shown in enclosure II.)

#### Policy Guidance for NAEP Under an Assessment System

Policy decisions concerning the role and nature of a national assessment (NAEP or some new version) will be critical to the design of a standards-based assessment system. Many important policy questions concerning NAEP as a key element in a larger system of standards-based assessments will need to be resolved as national content standards are certified. Such questions include:

- Should NAEP regularly assess every subject for which standards are established?<sup>16</sup>
- What portion of the content standards should NAEP cover?<sup>17</sup> Should it cover the standards as comprehensively as possible, within the limits of large-scale assessment technology (thus perhaps covering a "package" broader than that selected by many states)? Or should it focus on a core of skills and knowledge applicable to all states and all students (thus covering possibly a narrower range of performance than NAEP has covered before)?
- Will NAEP be expected to provide baseline data for evaluation purposes as well as to reflect the content standards? Or will more than one national test be needed?
- Given that NAEP will necessarily address only one testable "package" of material drawn from the standards, how should performance on NAEP be interpreted? Should there be national standards for performance on the national test?
- Should explicit linkages between NAEP and other state, regional, or commercial tests be constructed? If so, what safeguards against inappropriate comparisons and misuses of data should be established?

NAEP is the best source of national data about student progress in the subjects specified in the national education goals selected in 1989 at Charlottesville; thus, the Goals Panel as a whole will be vitally concerned with these questions, as will each of its constituent elements: the Congress, because changes in NAEP may require statutory authorization and budgetary support; the administration, because of the implications for the Department of Education; and governors and state legislators, because of concerns that any national test give due regard to the state responsibility for education and the voluntary nature of the national content standards.

The governance issue is: Will the certification of content standards call for the reassignment of policy responsibility for NAEP? We think that it will. The questions listed above amount

---

<sup>16</sup>Standards are currently being formulated with federal support for the subjects that are both mentioned in the national goals selected at the Charlottesville meeting and mandated for NAEP: mathematics, science, history, geography, and English (reading, writing, and speaking). Additional groups are working on standards for civics, social studies, the arts, and health education.

<sup>17</sup>The answer to this question may differ from subject to subject.

to a fundamental review of national testing and its purposes, at the high level of policy represented on the Goals Panel. To address them satisfactorily will require a combination of representation and negotiation (to work out the questions of state-federal relations) with technical expertise (to identify technically feasible options and to help negotiators understand the costs and the information benefits of each).

The Goals Panel is appropriately constituted to perform the negotiation and representation functions necessary to these policy decisions. Moreover, S.2 assigns the Goals Panel the functions of proposing indicators of progress toward the educational goals and benchmarks for evaluation. However, as noted earlier, the Goals Panel does not incorporate relevant technical expertise, nor does NESAC. An advisory mechanism that provides for a combination of technical and constituency input, and for the thorough and objective exploration of alternatives, will therefore be needed.

NAGB has begun to explore how NAEP's role might change under a system of standards and will be heard from as that exploration proceeds. NCES and the NAEP contractor will be essential sources of technical information. As parties at interest, however, neither NAGB nor NCES should direct the redefinition effort. Nor should NAGB give independent direction to NAEP concerning policy issues that have been resolved at the level of the Goals Panel.

#### The Issue of Authority for NAEP

The current governance structure does not make clear who can speak for NAEP in system-level discussions concerning its future or who will be accountable for its performance on matters concerning the standards and assessment system. NAEP governance is currently divided (see enclosures I and II): the Secretary of Education (who currently is one of the two executive branch representatives on the Goals Panel) bears administrative responsibility for the assessment, and NAGB bears responsibility for policy and for performance of specific quasi-operational functions. The issue of who has authority will need to be resolved.

#### NAGB's Specific Responsibilities

Whatever the structure for NAEP policy decisions, content standard certification and assessment review will have implications for NAGB's remaining responsibilities to determine the content of assessments, set achievement goals, and supervise the establishment of linkages between NAEP and other tests.

## Determining the Content of Each Assessment

NAGB now directs the work of "consensus groups" (comprised of educators and citizens) that plan the content to be covered on each NAEP test. Under a system of content standards, essential content for each subject and grade will already have been approved by a broadly representative body through procedures that reveal information about current practice as well as about desired improvements. Thus, much of the background work formerly performed by the consensus groups will have been accomplished. There will still be important NAEP assessment design decisions to be made.<sup>18</sup> But the assessment design groups working to incorporate national content standards into NAEP will face a different and more technical set of issues than previous consensus groups.

As noted earlier, the content standards are likely to specify some student capabilities that cannot be adequately measured by the techniques NAEP has typically used. Decisions concerning how NAEP can best respond to these standards will need the combined wisdom of educators, citizens, and experts in measurement. The current arrangements, in which test content is planned by educators and citizens, approved by NAGB on a policy basis, and then turned over to measurement experts for implementation are not likely to be adequate to the new demands. We conclude that revised arrangements that focus on implementing NAEP's role in the new system and foster the integration of technical with content considerations will be needed.

## Identifying Achievement Goals

The national content standards will identify content mastery goals, so it will not be necessary for NAEP to identify this type of "achievement goal." But the content standards will not identify goals for performance on the NAEP test, so the question remains whether NAGB should do so.

Since NAEP is the national test, any goal set for NAEP will take on the status of a "national goal" and will raise national and state issues similar to those raised by the content standards. We therefore consider the question of whether to set performance goals to be a higher level policy matter, and not one that should be left to NAGB.

If NAGB is to set performance goals, it should do so in a manner that is consistent with each certified set of content standards rather than independently deciding to report goals for all subjects in a certain way. Its task will be to find the way in

---

<sup>18</sup>For one thing, some time may elapse before standards for every subject assessed by NAEP are certified.

which NAEP data can best reflect each specific set of standards (a task that will require considerable technical expertise).

### Approving Linkages

Finally, the development of a system of assessments raises questions concerning NAGB's responsibility to oversee linkages between NAEP and other tests. NAGB currently sets linkage policy and delegates the task of technical review (which involves complicated statistical issues) to a Board committee and NCES. We have suggested that linkage policy be set as part of the larger process of assessment system design. If this change is implemented, such that the issues to be resolved are mainly technical, it is not clear what role will remain for NAGB.

### NCES Quality Control and NAEP Review Responsibilities

Currently, NCES is responsible for ensuring that NAEP data meet high standards of technical quality. This responsibility will continue. Under the proposed assessment system, technical quality and alignment will be reviewed externally for all tests that claim to measure performance in terms of the national content standards. We see no reason to exempt NAEP from this review, provided that it is conducted by a qualified body as discussed above.

NCES is also responsible for conducting reviews and validation studies of NAEP and of its usefulness. This function, too, continues to be needed and should be coordinated with the overall evaluation of the system of standards and assessments.

### CONCLUSIONS REGARDING GOVERNANCE ISSUES

We conclude that as the Goals Panel and NESAC assume their new duties and as content standards are certified, many of the representational and policy functions now assigned to NAGB will be performed, at least in part, at a higher level. The responsibilities that remain will be concerned primarily with implementing policy and will be more technical than before. Coordination, integration, and accountability, rather than independence, will be called for.

The context for NAEP will have shifted so substantially as to suggest a fundamental redesign of NAEP's governance. The redesign should emphasize responsiveness to high-level policy, unified direction and accountability, and integration of technical with subject-expert and constituent views at the level of NAEP operations.

We will send copies of this letter to interested parties upon request. If you have questions or would like additional information, please call me at (202) 512-2900 or Robert L. York,

Director of Program Evaluation in Human Services Areas, at (202) 512-5885, under whose direction this letter was prepared. Major contributors were Fritz Mulhauser, Assistant Director, and Gail MacColl, Project Manager.



Eleanor Chelimsky  
Assistant Comptroller General

**ENCLOSURES**

**I Groups Concerned With Standards and Assessments**

**II NAEP Governance Structure**

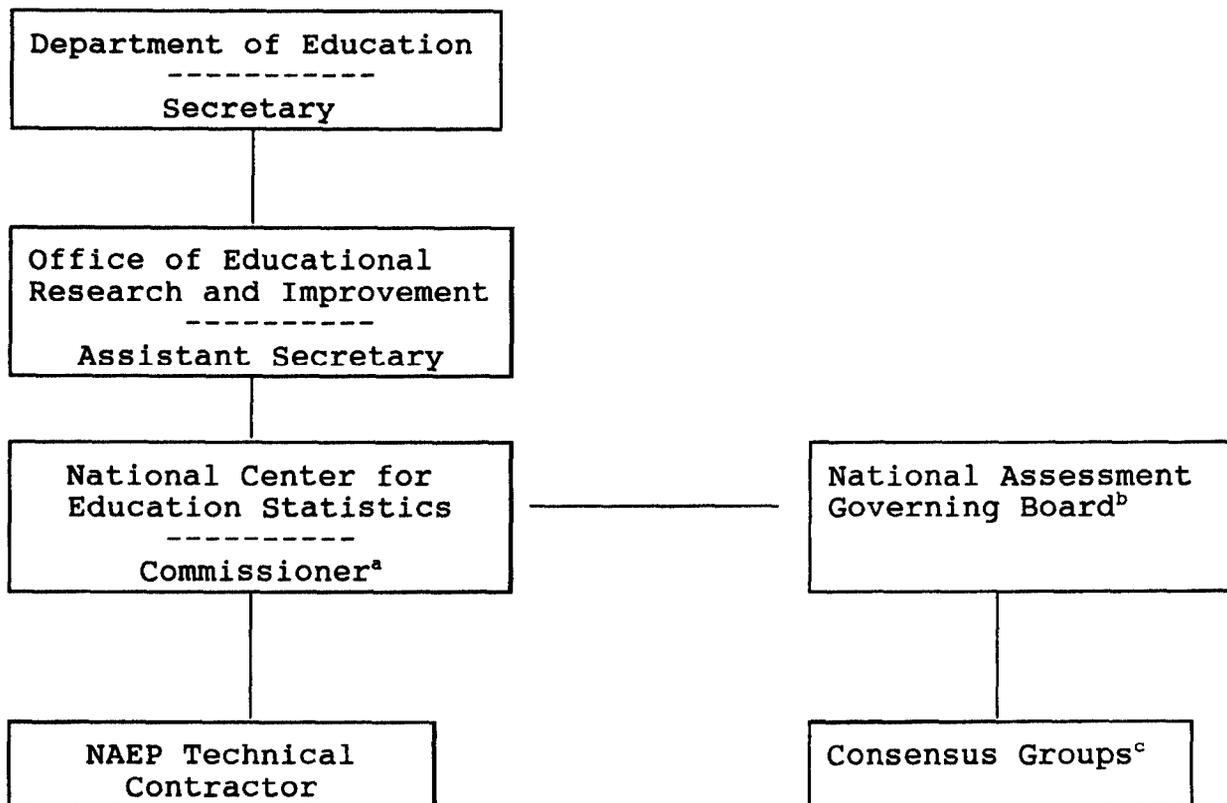
## GROUPS CONCERNED WITH STANDARDS AND ASSESSMENTS

Composition	Standards and assessment functions
<p>National Education Goals Panel</p> <ul style="list-style-type: none"> <li>2 Presidential appointees</li> <li>8 Governors</li> <li>4 Members of Congress</li> <li>4 State legislators</li> </ul>	<ul style="list-style-type: none"> <li>Establishes process for formulating standards</li> <li>Recommends groups to receive grants to formulate standards</li> <li>Certifies voluntary national content standards</li> <li>Certifies voluntary national school delivery standards</li> <li>Certifies assessments aligned to the content standards<sup>a</sup></li> <li>Publishes annual report of indicators of progress toward national education goals and recommending improvements in such data</li> <li>Periodically reviews and updates standards</li> </ul>
<p>National Education Standards and Assessments Council</p> <ul style="list-style-type: none"> <li>5 Educators (including elementary and secondary school educators)</li> <li>5 Members of the general public (such as public officials and representatives of businesses, foundations, and advocacy communities)</li> <li>5 Individuals with experience in educational assessment, content standards, and curriculum design</li> </ul>	<ul style="list-style-type: none"> <li>Develops criteria for reviewing voluntary national content standards</li> <li>Reviews and makes recommendations to the Goals Panel on voluntary national content standards and school delivery standards</li> <li>Reviews and makes recommendations to the Goals Panel about assessments aligned to these standards<sup>a</sup></li> <li>Develops criteria for model assessments of mathematics and science</li> </ul>

<p>School delivery standards development group</p> <p>Consortium of state officials, teachers, principals, school board members, representatives of concerned groups</p>	<p>Develops voluntary national school delivery standards and proposes them to the Goals Panel</p>
<p>National Academy of Sciences</p> <p>Technical experts</p>	<p>Recommends criteria for reviewing assessments</p> <p>Evaluates effectiveness of school delivery standards</p>
<p>National Assessment Governing Board</p> <p>2 Governors or former governors</p> <p>2 State legislators</p> <p>2 Chief state school officers</p> <p>1 School superintendent</p> <p>1 Member, state board of education</p> <p>1 Member, local board</p> <p>3 Classroom teachers</p> <p>1 Business representative</p> <p>2 Curriculum specialists</p> <p>2 Testing and measurement experts</p> <p>1 Nonpublic school administrator or policymaker</p> <p>2 School principals</p> <p>3 Representatives of the general public</p>	<p>Formulates policy guidelines for NAEP</p> <p>Selects subject areas</p> <p>Identifies achievement goals</p> <p>Develops assessment objectives and ensures that they were selected through a consensus process and that cognitive items are appropriate and free from bias</p> <p>Develops test specifications</p> <p>Designs assessment methodology</p> <p>Develops guidelines for analysis, reporting, and dissemination</p> <p>Develops standards and procedures for interstate, regional, and national comparisons</p> <p>Acts to improve the form and use of NAEP</p>
<p>National Center for Education Statistics</p> <p>Commissioner of Education Statistics</p> <p>Professional and technical staff</p>	<p>Implements NAEP through the services of a technical contractor</p> <p>Ensures fairness and technical quality of NAEP data</p> <p>Conducts reviews and validation studies of NAEP, solicits comments on its conduct and usefulness</p>

\*Possible addition to currently specified functions.

## NAEP GOVERNANCE STRUCTURE



<sup>a</sup>The Commissioner of Education Statistics administers NAEP with advice from NAGB and reports to NAGB (on behalf of the Secretary) concerning the department's actions to implement the board's decisions.

<sup>b</sup>The National Assessment Governing Board provides policy guidance for NAEP. Members of NAGB are appointed by the Secretary of Education, but NAGB is independent of the Secretary and the Department of Education.

<sup>c</sup>Consensus groups of educators and citizens plan the content to be covered in each NAEP test.

(973774)