THE NATION'S REPORT CARD

Improving the Assessment of Student Achievement

Report of the Study Group
Lamar Alexander, Governor of Tennessee, Chairman

H. Thomas James,
Vice-Chairman and Study Director

With a review of the report by a Committee of the National Academy of Education

Robert Glaser, University of Pittsburgh,
Chairman

THE NATION'S REPORT CARD

THE NATION'S REPORT CARD

Improving the Assessment of Student Achievement

Report of the Study Group
Lamar Alexander, Governor of Tennessee
Chairman

H. Thomas James
Vice-Chairman and Study Director

With a Review of the Report by a
Committee of the National Academy of Education

Robert Glaser, University of Pittsburgh
Chairman

The National Academy of Education

The National Academy of Education is composed of scholars and distinguished professionals, all of whom have been elected to the Academy in recognition of their outstanding contributions to education. The Academy has a long-standing commitment to foster public understanding of education and to provide comment on and analysis of educational issues to the government. In the spring of 1986, the Department of Education requested an Academy review of a report on the National Assessment of Educational Progress that a panel convened by Lamar Alexander and H. Thomas James was to prepare. The National Academy of Education agreed to participate in this effort. Specifically, its role included three activities:

1. Appointment of a small, independent committee of its own to comment upon issues of assessment, in the context of the Alexander-James Study Group's report;
2. Coordination of publication and distribution of the reports of the Alexander-James Study Group and the Academy Committee;
3. Coordination of fiscal administration for the entire project.

We are pleased to present the report of the Alexander-James panel and the independent commentary by a committee of the Academy. We believe that these documents taken together make an important contribution to current discussions about the National Assessment of Educational Progress and to its role in improving our understanding of educational achievements in the United States.

Patricia Albjerg Graham
President, National Academy of Education

Copies of this report can be obtained by sending your order with $9.00 for each copy to:
National Academy of Education
Harvard Graduate School of Education
108 Longfellow Hall
Cambridge, MA 02138

Contents

January 15, 1987

The Honorable William J. Bennett
Secretary of Education
Federal Office Building
400 Maryland Avenue, S.W.
Washington, D.C. 20202

Dear Mr. Secretary,

I am pleased to transmit to you the report of the Study Group on updating the nation's report card.

Parents, policymakers, professionals, and other taxpayers need timely, factual information on what students know and can do. This is essential if our nation is to plan well and achieve its goals of better schools, better jobs, and a stronger people.

The existing national report card—in the form of the National Assessment of Educational Progress—has helped fill that need, but it must begin to go even further.

The continuing wave of school reform across our nation makes this need ever more urgent. Communities, schools, school districts, and states are in the midst of momentous decisions that will affect the quality of American education.

Decision makers need to see the facts clearly. They must make sense of a storm of confusing data and help lead the way to better schools. The nation's report card—if it is well-designed, clear, and usable—can be a rudder against the storm. Our report shows how that can be accomplished.

To produce these findings and recommendations, the Study Group has reviewed the history of NAEP, examined its accomplishments and shortcomings, and sought the advice of some of our nation's most respected experts in the assessment field. We hope our conclusions are helpful as you plan your next steps.

Yours sincerely,

Lamar Alexander
Chairman

Acknowledgments

This report on how we assess what students know and can do, and the proposals for improving that assessment, was made possible by funds provided by the Department of Education, the Exxon Education Foundation, the Ford Foundation, the William and Flora Hewlett Foundation, the John D. and Catherine T. MacArthur Foundation, and the Matsushita Foundation. The National Academy of Education coordinated the fiscal administration of foundation funds, and a committee of the academy chaired by Robert Glaser reviewed this report. Their separate views are a part of this document. We are grateful to Patricia Albjerg Graham, president of the academy, for her interest and helpful counsel and to Gail Keeley, executive director of the academy, for her unfailing good humor and efficiency. We are also grateful to Chester E. Finn, Jr., Emerson J. Elliott, and David Sweet, whose close attention to the complex logistical liaison kept our operation on what seemed at first an impossible schedule.

The most important contribution to our enlightenment about the complex nature of the assessment came to us through the forty-six commissioned papers. Special thanks are due to the authors who had to forgo other opportunities, including in many instances vacations, to complete the papers commissioned in May and June of 1986 in time for our September 30 meeting. Authors and titles of these papers (accessible through the ERIC system) are listed in Appendix B of this report. Also listed (Appendix A) are the names of the conveners and members of the nine topically organized subgroups that met in various parts of the country during the spring and summer, originally to help us select authors and subjects of papers to be commissioned, and later to condense the findings in ways best suited to informing us on possible courses of action.

The Council of Chief State School Officers will be especially interested in our report because of that organization's work in planning for state assessments. We are grateful to Richard A. Boyd for his involvement in our deliberations and to Ramsey Seldon for his work with several of our subgroups and for sharing with us materials that helped us understand the plans for the State Education Assessment Center he directs.

We were informed in the beginning that we had no responsibility for reviewing the performance of the grantee currently administering the National Assessment of Educational Progress. We are especially grateful

to Gregory Anrig, Archie Lapointe, and Paul Barton, of Educational Testing Service, for their extraordinary effort to provide the Study Group, its staff, the conveners of the subgroups, and the authors of the commissioned papers with a remarkably rich flow of information.

We appreciate the thoughtful and helpful comments of officials of a number of educational associations. These include Samuel G. Sava of the National Association of Elementary School Principals, Scott D. Thomson of the National Association of Secondary School Principals, Gordon Cawelti of the Association for Supervision and Curriculum Development, Robert L. Smith of the Council for American Private Education, John C. Esty, Jr., of the National Association of Independent Schools (who also served on our Study Group), Thomas A. Shannon of the National School Boards Association, and the Reverend J. Stephen O'Brien of the National Catholic Educational Association.

Finally, the dedication and hard work of our staff was a crucial ingredient in getting this report together: Annegret Harnischfeger, deputy director, Thomas B. Hoffer, research associate, and assistants Mary E. Driscoll, Suzanne Gaskins, Michael Shirley, and Samuel P. Whalen. Joan Masters managed our clerical services and kept lines of communication open.

Space does not permit acknowledging all of the colleagues and friends who were generous of their time in offering suggestions and responding to questions; a partial list would include Joy McLarty, who attended all of our meetings, Arthur G. Powell, Senta A. Raizen, Gail E. Thomas, James S. Coleman, Peter Gerber, Robert Linn, Michael A. Olivas, George W. Neill, David Hornbeck, Keel Hunt, and Bruce Thomas.

A large and very special debt of gratitude is owed to Ralph Tyler, who prepared one of the commissioned papers, attended every meeting of the Study Group, and gave generously of his time in responding to questions and offering thoughtful advice throughout the study period. His role in our efforts is unique, not only as a "founding father" of the National Assessment of Educational Progress more than a quarter of a century ago, but as a key participant in all its later history.

Part 1

Report of the Study Group

The Study Group

Lamar Alexander, Governor of Tennesee and Chairman,
National Governors Association
Chairman

H. Thomas James, President Emeritus, the Spencer Foundation
Vice-Chairman and Study Director

John K. Andrews, Jr., president, Independence Institute

John E. Brandl, professor of public affairs, University of Minnesota, and member of the Minnesota House of Representatives

Owen B. Butler, former chairman of the board, Proctor and Gamble, and chairman, Subcommittee on Business and the Schools, Committee for Economic Development

John T. Casteen III, president, University of Connecticut, Storrs

Hillary Rodham Clinton, partner in Rose Law Firm, Little Rock, and First Lady of Arkansas

Antonia Cortese, first vice-president, New York State United Teachers, and vice president, American Federation of Teachers

Linda Darling-Hammond, director, Education and Human Resources Program, Rand Corporation

John Ellis, superintendent of schools, Austin (Texas) Independent School District

John C. Esty, Jr., president, National Association of Independent Schools

Pascal D. Forgione, Jr., chief, Office of Research and Evaluation, Connecticut State Department of Education

Richard A. Giesen, chairman of the board, American Appraisal Associates

Larry V. Hedges, associate professor of educational statistics, University of Chicago

Oscar G. Hernandez, president, San Antonio Independent School District, and president, Texas Association of School Boards

Bill Honig, state superintendent of public instruction, California State Department of Education

Lyle V. Jones, director, L. L. Thurstone Psychometric Laboratory, University of North Carolina at Chapel Hill

Francis Keppel, senior lecturer on education, Graduate School of Education, Harvard University

Michael W. Kirst, professor of education, Stanford University

Evelyn F. Luckey, assistant superintendent of elementary schools, Columbus (Ohio) Public Schools

David G. Savage, *Los Angeles Times*

Jacqueline H. Simmons, principal, Paul Robeson High School, Chicago

Introduction

How well are our children doing in school?

What are they learning?

Are our schools doing a good job?

Is our education better than it used to be?

Will students in our state, and in our nation, be able to compete successfully in the future?

Americans ask questions like these every day, but the answers to them, and to other important questions concerning the adequacy of our educational system, are hard to come by. It is not that we lack the ability or the will to measure our students' progress. For many decades we have been studying American schooling, officially and unofficially, spending many millions of dollars in the process. And with every fresh effort we learn things about our schools and our students that we did not know before. But for a number of reasons, some good and some bad, some historical and some technical, current assessments are not producing answers to the questions most often asked at this moment in history by parents, by concerned citizens, and by educators.

It was for these reasons that Secretary of Education William J. Bennett formed our Study Group in May 1986. Our assignment was to take a fresh look at how we in the United States currently assess what our students across the country know and can do, and to suggest, if we could, ways of improving the process.

We began by examining the ongoing National Assessment of Educational Progress (NAEP). This federally funded project, initiated more than twenty years ago, tests a nationally representative sample of nine-, thirteen-, and seventeen-year-olds to see how they are performing in reading, writing, science, mathematics, and other important areas. The "nation's report card," as NAEP is now often called, is the most prestigious, most extensive, and most valuable of our assessment efforts, and the high-quality data it gathers and publishes are widely used in research, planning, and school administration.

GETTING MORE HELP FROM NAEP

But NAEP has a serious weakness, and this must be identified here at the outset, for correcting it is our Study Group's most important recommendation. The weakness is that while providing excellent information on what our children know and can do, it provides it only for the nation *as a whole*, and for a few large regions of the country. Whole-nation information is of course useful when we wish to gauge the performance of our children against that of children in other countries, whether rivals or allies. But in the United States education is a *state* responsibility, and it is against the performance of children closer to home that we want and need to compare the performance of our youngsters.

Nearly all important decisions in education are made at the state and local levels, and accountability for performance is vested at those levels. Yet rarely do the governors, legislatures, chief state school officers, boards of education, and other decision makers—let alone the parents of school-age children—have all the information they need on how well their schools are doing.

Certainly, local schools and districts assess their students' progress, and they have information that is useful in guiding individual students through their educational programs. Many school districts, and some states, use nationally standardized tests to compare the performance of their students with a reference group of students nationwide. These efforts provide useful information, but rarely is that information coordinated and organized so that it is possible to make responsible comparisons over longer periods of time or between states. If the governor of Tennessee wants to know whether the school children of his state are better readers than they were twenty years ago, or better writers than their counterparts in Texas, North Carolina, or Indiana—and these are just the sorts of things that the chairman of this Study Group has wanted to know during his time as governor of Tennessee—it is practically impossible for him to find out.

Nor is it a simple matter for educators and policymakers to get reliable current data on what some subgroups of our young people know and can do—though such information is crucial if we are to address concerns about equality of opportunity, children "at risk," and those with special needs. It is especially difficult to get information on youngsters who do not attend our public schools, as for example, our "dropouts."

Of course, no single report card can supply every bit of information that everybody wants. This is a big and diverse nation; different people naturally want to know different things about its education system, and to know them at a half a dozen "levels," from the individual child through

the classroom, the school, the local education system, and the state, to the country as a whole. Although the national assessment cannot supply all these wants directly, it can, we are convinced, supply information at lower levels, to decision makers in smaller jurisdictions, than it does at present. And this will be to the enhancement of education throughout the nation.

If we think of NAEP as a weather map, today's assessment is designed to provide temperature, barometric pressure, and precipitation levels only for the United States as a whole and for a few large regions within it (the Midwest, for example), regions that are essentially meaningless for education policy matters. We propose, instead, a much expanded weather map that will not only provide such information for the whole country, but will also provide it for every state within it—and do so in such a way that a state or locality can readily produce similar data at the community or even neighborhood level. These data in turn can be compared with data from other communities, the entire state, or the nation, both now and over time.

PREVIOUS CONCERNS ABOUT COMPARISONS

In earlier years, some educators resisted gathering this kind of information. They did not think that comparisons were necessary or fair, nor did they welcome the accountability that such an information base makes possible. Some felt that the "intangibles" of education—those qualities of character and spirit that testing programs cannot readily gauge—were so important that to gather information only about skills and knowledge that are most easily measured might distort the education process itself. Others felt that since youngsters should not all be held to the same standard of performance, they should not all be subject to the same assessment procedures. And there has long been a concern that any sort of nationwide assessment program would somehow generate a single national curriculum for all schools and all schoolchildren—something that many Americans would find objectionable.

Our Study Group concluded that many of these concerns are less important now than they were previously, and that most can be readily accommodated within a redesigned national assessment. We also concluded that the rewards will be fully worth the effort. Adoption of our recommendations will vastly improve our ability to keep track of what our children know and can do. Properly structured, accurately described, and carefully delimited, these changes will give rise to no serious new problems. Moreover, the governance system we propose for the nation's report card will be highly responsive to public and professional con-

cerns, while staying clear of inappropriate political influence.

A big and very significant change has come over American education in the recent past. Today, rather than wanting to conceal, suppress, or avoid knowledge of how the educational progress of children in one state compares with that in another—or with the average for the nation as a whole—we now accept such comparisons as legitimate and desirable. In response to this new climate in the world of education, the federal government for the past several years has provided state-specific education performance data in the form of the Secretary of Education's annual "wall chart." The problem with the wall chart is that its main sources of performance or "outcomes" information are wretchedly inadequate for the purpose. The wall chart relies on state average scores on the two big college entrance tests that many high school seniors choose to take. We know why the Education Department uses these data: there are simply no other comparable data available. But just about everyone agrees that there ought to be. And our Study Group has concluded that a reconstituted national assessment should provide them.

HOW NAEP WORKS AT PRESENT

In order to understand how the nation's report card that we recommend will differ from—and be more informative than—the present approach, readers may benefit from a sketch of NAEP as it exists today.

The National Assessment of Educational Progress is a program authorized by Congress and administered by the Department of Education,* which functions by means of a grant to a private nonprofit organization (currently the Educational Testing Service, previously the Education Commission of the States). The basic federal payment was approximately $4 million in fiscal 1986. Tests are administered every two years to a nationally representative sample of youngsters at the ages of nine, thirteen, and seventeen. Most of these boys and girls are enrolled in grades 3, 7, and 11.

The current law requires that reading, writing, and mathematics be assessed every five years. Decisions about other subjects to be assessed, the frequency with which this is to be done, and other major policy decisions concerning the governance of NAEP are made by an Assessment Policy Committee consisting of educators and laymen chosen by the grantee.

In addition to knowledge-and-skill tests, the youngsters in the NAEP sample provide background data about themselves, their families, and their schools. Some of the teachers and principals in the same

*See Appendix C for the enabling statute.

schools also supply information on the instructional process for subject areas being assessed and on the school.

In an ordinary test cycle, the NAEP sample includes approximately twenty-five thousand students at each of the three age levels. This is enough to give reliable data about performance levels for the nation as a whole and for four geographic regions within it; for boys and girls; for white, black, and Hispanic youngsters; and for children living in several types of communities.

NAEP does not supply any state-specific data, although states (or others) may purchase the added testing that will provide information about student performance in some or all of the subject areas and age levels. To date, fourteen states have participated in some aspect of this, although their specific arrangements have differed widely.

FOR THE FUTURE: OUR RECOMMENDATIONS

In the Study Group's work on the national assessment we have reached a number of important conclusions about the functioning of an improved NAEP. Seven key recommendations are summarized below and discussed more fully in the chapters that follow this Introduction. In all our work we have been guided by the overriding principles of making the nation's report card more useful to states and localities, assuring its responsiveness to state and local concerns, shielding it from partisan politics, and improving the quality, timeliness, and comprehensiveness of the information it provides. Our work has been vastly aided by the expertise and hard labors of nine subgroups and informed by the wisdom distilled in some forty-six papers commissioned by our main group.*

1. Maintain Continuity

The information that NAEP has generated since 1969 is the best "baseline" data available concerning what children know and can do. As the nation's report card is improved, it should maintain continuity with this fine data base. Yet the design of NAEP should take full advantage of state-of-the-art advances in testing, sampling, computer-adaptive procedures, and analysis. It should also realize the full potential of NAEP as a rich source of descriptive information that can put achievement findings into the context of school instructional processes and student background.

*The members of the subgroups are identified in Appendix A of this report and the commissioned papers listed in Appendix B.

2. Assess the Core Curriculum

The assessment should cover those subjects and skills that are usually taught in schools and school systems; that are generally regarded as central to the education of American youngsters; and that lend themselves to testing across the breadth of the nation. We urge regular assessment of reading, writing, and literacy; mathematics, science, and technology; and history, geography, and civics. Other skills and subject domains should from time to time be included. In every instance, the assessment instruments should examine acquisition of pertinent "higher-order" skills as well as basic skills, knowledge, and concepts. And in every instance, the objectives for content areas and the reporting and interpretation of results should be clear and meaningful to educators, policymakers, researchers, and the general public.

3. Focus on Transition Grades and Expand the Sample

As in the past, the nation's report card should continue to gather information on children aged nine, thirteen, and seventeen, but grade-level samples should be changed from the present grades 3, 7, and 11 to the more important "transition" grades of 4, 8, and 12. In addition, out-of-school seventeen-year-olds should be included and, in the assessment of literacy, older age groups should be included as well. By making these changes, we will regularly gather vital data about two of the most important issues in American education today: dropouts and adult literacy.

4. Create an Educational Assessment Council

The governance and policy direction of the national assessment should be furnished by a broadly representative Educational Assessment Council that provides wisdom, stability, and continuity; that is charged with meshing the assessment needs of states and localities with those of the nation; that is accountable to the public—and to the federal government—for stewardship of this important activity; but that is itself buffered from manipulation by any individual, level of government, or special interest within the field of education. The broad categories of membership of the Educational Assessment Council are described in the report and

should be mandated by statute. The Secretary of Education should appoint the members—for overlapping five-year terms—but would be obliged to select from among candidates recommended to him by a permanent, statutory nominating committee.

A separate test contractor under contract with the federal government should handle test development, administration, analysis, reporting, maintenance of item banks, and provision of assistance to states and others in supplementing tests. It would be guided by policies established by the council concerning test domains, learning objectives, test design, and plans for analysis.

Thus the overall governance of the nation's report card would consist of three major elements, each with specific duties, powers, and rights: the Educational Assessment Council, the testing contractor, and the federal government. This structure is meant to supply needed checks and balances and "separations of power" for this important and sensitive enterprise.

5. Provide for Add-on Assessments

The basic report card for the nation should be based on a statistical sample that is accurate at the state level but that preserves the anonymity of districts, buildings, classrooms, and individual students. But it will be possible for states and localities to augment the national assessment—and such augmentation will be encouraged—so that information concerning the knowledge and skills of children at the district and even the school-building level can be made available.

6. Assess Private School Students

As has always been the case, private school students would be included in the sample along with public school students, but the sample should be large enough to draw valid conclusions about student achievement in individual grades and major subgroups of private schools. And, just as a state or locality may augment the assessment to get more detailed information about its students, so also could an organization of private schools enlarge its portion of the sample in order to obtain additional information about the overall knowledge and skills of youngsters attending its schools.

7. Fund the Essential Assessment

The federal government must assume responsibility for meeting the core operating costs of a high-quality report card on the condition of American education. This will require a change in funding from the four- to six-million dollar level that has characterized NAEP in recent years to approximately five times that much. But it is essential to see this in perspective: the $20 million to $30 million a year that we estimate it will cost to produce a first-class report card is a minuscule sum within an elementary and secondary education system that is spending approximately $170 billion this year, and for a Department of Education that now spends approximately $20 billion on its various programs. We are recommending a price tag that is equivalent to about fifteen cents out of every thousand dollars that the United States spends on schooling; a tiny expenditure for finding out how good that schooling is. In terms of the federal Department of Education budget, what we are recommending is that slightly more than one tenth of one percent of the agency's current appropriation be devoted to this central activity, which will benefit the entire nation. States, localities, and others would of course be expected to pay for any additions or augmentations that they may ask for.

IN CONCLUSION

Our Study Group has completed its assignment, but the hard work remains. The issues addressed in the preceding paragraphs and the following pages are as urgent and consequential as any facing American education in the late eighties. There is pressing need for a new consensus about the principles, goals, and practices that will guide how we determine what our children know and can do as they pass through our schools and into adult society.

Our recommendations offer a sound basis for that consensus, one that necessarily and properly goes far beyond the federal government. The need for a well-designed, clear, informative, and functional report card on our schools and our children is felt keenly by states, local school districts, education professionals, parents, and the general public. They must be part of the new consensus about how to meet that need.

Designing and carrying out broad-based education assessments takes time. Years pass between the conceptualization of an assessment and the earliest returns from it. It is therefore all the more urgent that the central principles governing the future of the National Assessment of Educational Progress be adopted as soon as possible and that the resources for carrying them out be made available promptly.

1

Roles and Responsibilities

The National Assessment of Educational Progress has become the primary source of information on what American students, nationwide, know and can do, covering a wide range of school subjects, with primary emphasis on reading, writing, and mathematics. Since by design that information is comparable in many respects from one assessment to the next, we can now compare the children of today with those of earlier years and learn how well our educational system is working. No other measure of educational achievement has NAEP's size, scope, and perspective over time.

Not only has the assessment yielded broadly based information about children's academic achievement, it has also established a standard of excellence and innovation for educational assessment in general. The scholars and educators who pioneered the original NAEP insisted from the outset that the information gathered be of the highest quality that current technology and theory would permit. Since that time the assessment's growing influence and reputation bear witness to NAEP's fulfillment of its founders' intentions.

We hasten to point out, however, that the original design of the national assessment was conceived in a far different climate of opinion from that obtaining today. During the 1960s school systems and professional educators generally opposed comparisons of student achievement among states, districts, or individual schools. The consensus at that time was that such comparisons could not be made fairly, since student backgrounds and school resources varied so widely. Consequently, the assessment was designed to limit the level of comparison to four geographical regions of the country and to certain subgroups of students.

STATE-BY-STATE COMPARISONS

The single most important change recommended by the Study Group is that the assessment collect representative data on achievement in each of the fifty states and the District of Columbia. Today state and local school administrators are encountering a rising public demand for thorough information on the quality of their schools, allowing comparison with data from other states and districts and with their own historical

records. Responding to calls for greater accountability and for substantive school improvements, state officials have increasingly turned to the national assessment for assistance.

In the past many states have participated in the assessment process and have received assistance in devising their own assessment programs. Other states have opted to purchase "add-ons" to the national assessment in order to obtain representative information for their own states. But until recently, little effort has been made to coordinate assessment methods employed at the national, state, and local levels. The changes we recommend would bring substantial benefits to each of the fifty-one governmental units, while preserving the national record and maintaining the educational autonomy of states and localities. All states would receive a core of information collected through a centralized mechanism, and would thus be guaranteed high-quality, fully comparable data. At the same time, individual states can supplement this core with additional data collections of their own. The national sample will continue to provide information about the achievement of policy-relevant subgroups and about the processes responsible for differences in achievement.

The Study Group further recommends that the national assessment provide state-by-state comparisons of core content areas (reading, writing, and literacy; and mathematics, science, and technology; history, geography, and civics). The redesigned assessment will yield average scores in these core areas for the nation as a whole and for each state. In addition, NAEP should continue its periodic measurement of student performances in a number of other content areas.

These purposes were not those to which the original design of the national assessment was addressed. To address them now, major changes in its design will be required, as well as in the dissemination of its results. The Study Group envisions the future national assessment as an integrated data-collection effort, providing policy-relevant information to decision makers at the federal, state, and local levels. The national assessment can become this if its current stewards act with the same imagination and foresight that marked the work of the assessment's founders.

LINKAGES WITH OTHER ASSESSMENTS

Several states and local school districts have instituted their own assessment programs in recent years and will want to continue these efforts alongside the expanded national assessment. Recent developments in test theory and measurement technology now make it possible to compare scores from different assessment instruments, thus broadening the scope of comparisons that can be made. We recommend that the national assessment devise a linkage system relating local and state testing and assessment programs to the national assessment. This will require considerable conceptual as well as technical work, for it remains an imperative that the curricular autonomy of the states and localities be respected and fostered.

Recent years have also witnessed an increasing interest in the use of national assessment data for international comparisons of student performance. The International Association for the Evaluation of Educational Achievement (IEA) has been sponsoring international studies of educational achievement for more than twenty years. Uncertainty about the commitment of the United States to this project, however, has often hindered funding arrangements and eroded the quality of United States study results.

Discussions are now underway to establish a U.S. Consortium on Cross-Cultural Studies of Education which would expedite and oversee our future participation in international assessments. In light of the growing interest in international comparisons, we recommend that the agencies involved in these discussions proceed with all due speed to establish and empower this organization. Since the national assessment is fast becoming the recognized standard for American achievement studies, we further recommend that the new consortium investigate ways of relating to the national assessment.

A NEW GOVERNANCE FOR NAEP

In order to undertake such demanding new tasks as state-by-state comparisons, the national assessment will require some important changes in its current governance structure. State and local school systems vary widely in their specific curricular goals and their preferred methods of assessment. As a means of arriving at a broadly based consensus regarding the aims and objectives of national assessment, we recommend the creation of an independent governing agency, the Educational Assessment Council (EAC). The EAC would operate independently of the institution carrying out the actual assessment and would define content areas, assessment procedures, and guidelines for fair comparisons of states and localities. The council would also maintain regular consultations with the proposed Consortium on Cross-National

Studies in order to coordinate the activities of the two agencies. We outline the constitution and proposed functions of the EAC in more detail in chapter 4 of this report, entitled Governance.

2

Content and Coverage

Federal statute requires the national assessment to measure student achievement in the areas of reading, writing, and mathematics on a regular cycle, and to report periodically on student knowledge and skill levels. NAEP is directed also to undertake special assessments in response to education's changing informational needs. While testing reading, writing, and mathematics on a regular schedule since 1969, NAEP has assessed achievement in a number of other areas, such as science, computer competence, music, art, and citizenship. (A summary of past assessments may be found in Appendix D of this report.) In what follows, the Study Group recommends the measurement of core content areas on a regular cycle: reading, writing, and literacy; mathematics, science, and technology; and history, geography, and civics. Beyond these, the national assessment should continue to undertake special assessments.

One theme that will continually reappear in this report is a concern for the measurement of more complex levels of thinking and reasoning. We are convinced that it is time for the national assessment to devote closer attention to the measurement of more complex skills.

HIGHER-ORDER THINKING SKILLS

From its inception, NAEP has devoted attention to the assessment of higher-order thinking skills. Currently a pilot study addressing complex thinking in mathematics and science is underway. At the highest level these skills include recognizing a problem's general structure, defining goals, isolating the information relevant to problem solutions, and evaluating the merits of arguments. Other high-level skills include reasoning, analyzing, explaining, and finding analogies. These skills enable a student to organize, coordinate, and direct lower-level skills effectively.

Until recently, however, careful definition and measurement of these skills has taken second place to a more limited measurement of factual knowledge based largely upon memory or very simple reasoning tasks. The Study Group strongly recommends that higher-order thinking be made a primary concern of future assessments of national achievement. In this we concur in the unanimous consensus of the subgroup exam-

ining the assessment of higher-order learning: "The measurement of higher-order cognitive skills is essential if we are to determine the extent to which the goals of education are actually being met" (Miller 1986).

We recognize of course that the effort to define, assess, and teach this level of skill is a well-established one in education. Education at its best has always stressed the value of learning how to think, and not merely knowing what to think. What is new, however, is the emergence of a commitment in recent years to develop and assess these skills among all members of the national community. Such a challenging objective had to await the development of new insights into the exact nature of these skills, and the best ways to measure and evaluate them.

Recent developments in the many allied disciplines that study thinking and learning have now placed the accomplishment of this goal within our reach. The past two decades have witnessed an explosion of new ideas about how human beings learn and what happens in higher-order thinking. At the same time powerful new computer technologies now make available a range of measurement tools sophisticated and sensitive enough to measure complex thinking with some accuracy. These new ideas and methods suggest fresh approaches to familiar old problems that have hampered the definition and measurement of higher-level thinking in the past.

The national assessment needs to identify the specific kinds of higher-order processes that organize and control simpler levels of skill in such varied subject areas as mathematics, science, writing, and reading. In the same way the assessment should be careful to identify which kinds of higher-order tasks are appropriate to each age level. The newest developments in research on learning imply above all that children can begin to acquire these thinking and learning skills at an earlier age than previously expected. That is, the development of skilled and flexible thinking does not need to wait upon the mastery of more "basic" or "fundamental" skills grounded in rote memorization. Recent evidence indicates that young children are able to bring some of these higher processes of thought to problem solving when the tasks do not place too heavy demands upon their more limited memory skills. Research findings also show that not all subject areas require the same types of thinking skills.

The national assessment should also use new measurement technologies to develop assessment methods that go beyond the limitations of the standard multiple-choice format. Multiple-choice examinations may be easier to score and more economical to administer. But they do not easily highlight and measure those higher-order skills that lead up to and organize simpler skills like computation or the use of memory. "Methods

of testing are needed that will probe beyond memory into those higher cognitive processes that the schools are expected to strengthen. To develop such methods, a clear conception is needed of what those higher-order cognitive skills are and how they can be measured" (Miller, p.4).

We urge NAEP to build upon the recent developments in cognitive research and measurement technology, as well as its own special strength as an assessment program to explore more effective strategies for identifying and measuring higher-order thinking. Unlike other indexes of student achievement, NAEP has always measured student performances by first defining a number of general learning objectives. Each objective in turn contains a scale of items that vary in their degrees of difficulty. As a result the assessment has amassed a wealth of knowledge about skill-level differences in subject areas like mathematics, science, reading, and writing. In addition, the national assessment has experimented in the past with methods of evaluation other than the standard multiple-choice format and has indeed long been recognized as an innovator in the field of achievement evaluation.

The Study Group therefore makes the following recommendations. First, we urge NAEP to make the evaluation of higher-order thinking a central concern of future assessments. It is essential that NAEP channel additional resources into developing ways to identify and measure the higher processes of thinking and learning. Second, NAEP should take special pains to investigate higher-order thinking in the critical subject areas of mathematics, science, reading, and writing.

Finally, we realize that genuinely effective assessment of complex thinking is an expensive proposition and will require special funding. But we are convinced that this additional expenditure is of great importance to the accomplishment of our nation's long-term educational goals. We recommend the appropriation of adequate funds for this purpose.

READING, WRITING, AND LITERACY

Reading and writing skills have been a regular component of national assessment since its inception, and the results of these assessments have proved useful to policy makers and practitioners in a wide range of educational contexts. Reading and writing skills in turn constitute the basis for literacy. There has been one full assessment of literacy to date, in 1985.

Reading

The national assessment has provided valuable information about students' reading abilities, including whether they can comprehend, analyze, interpret, and evaluate what they read. It has also measured students' attitudes about and interest in reading. There are, however, aspects of reading assessment about which additional information would be valuable. These include clarification of objectives and exercises, specific assessment of skill levels, information about schools and teachers relevant to reading, and assessment of reading readiness.

Over the years, the objectives of reading assessments have been broadened in important ways. We recognize that a consensus in defining objectives in the field of reading is particularly difficult to achieve. Thus it is important that the process by which objectives and exercises are developed continue to involve a broad range of experts and informed citizens. Such a process of consensus will insure that the resulting objectives are well understood and are an accurate reflection of the best current conceptions about reading. This concern became especially relevant with the recent development of a reading proficiency scale. This scale is more abstract than previous measures of reading outcomes, and less obviously tied to the objectives and the exercises. Those who use it will need considerable information about the exercises (beyond the "benchmark items" given for the 1984 assessment) to allow them to interpret the scores meaningfully.

We would also point out that student failures in complex reading tasks do not necessarily stem from deficiencies in higher-order reading skills. Success in such tasks clearly depends on the firm establishment of lower-order skills as well. The assessment should include explicit evaluation of both lower-level skills (such as knowledge of word meanings and sentence understanding) and higher-level skills (such as inference making and the drawing of conclusions) at all age levels. Assessment results should point out differences between these two levels of skill development. Insights into the sources of student reading performance would help educators and researchers to evaluate the need for particular instructional approaches. Information about schools and teachers which is directly related to reading also merits additional attention.

Reading could also be assessed in young children to yield insights into the early stages of reading and to explore connections between early and later reading success. We agree that these would be desirable outcomes of earlier reading assessment. Measurement of reading before age nine, however, is not feasible for a large-scale, nationwide assessment, since instructional approaches to early reading vary so widely.

Some of the advantages of measuring reading at an earlier age could be realized by including in the assessment a measure of reading readiness for children who are just beginning to learn to read. We recommend that the assessment explore such a measure.

Writing

Beyond accurate measurement of student achievement in writing, the purpose of the writing assessment is to assist schools in identifying objectives, tasks, and measurement techniques which could be used to improve the quality of their students' performance. Currently, the national assessment covers five major writing objectives: writing as a way of thinking and learning, writing to accomplish a variety of purposes, managing the writing process, controlling forms of written language, and appreciating the value of writing. Achievement within these objectives is measured by performance on three types of writing tasks—informative, persuasive, and imaginative—and by responses to questionnaires.

We urge the national assessment to develop and articulate a more coherent framework for the measurement of writing. This framework should identify common writing objectives, the best methods for measuring writing skills, and meaningful ways of scoring, interpreting, and reporting results. The assessment should also clarify the relative importance of different skills for writing outcomes. Of special concern are questions about why some skills and methods are selected for measurement over others, and how these particular skills are pertinent to those required in school and the workplace. Whenever possible, the national assessment should provide specific and age-appropriate examples of these skills to guide our understanding of what students know and can do at different points in their development.

The national assessment might further improve its writing report card by supplementing its current measurement practices with other methods, such as untimed writing exercises and a wider choice of topics. The scoring system should be comprehensible to teachers and should employ both general-impression scoring and well-defined ratings of desired features. The national assessment should draw more on the expertise of educators who are familiar both with students at each assessed age/grade level and with the relevant content. The assessment's writing report card should give specific examples of items and responses demonstrating particular skills. Finally, it is important that the report card provide simple and direct accounts of its findings and procedures to all interested audiences.

Literacy

The skills and knowledge encompassed by the concept of literacy overlap to a considerable extent the skills and knowledge currently tested in the reading, writing, and mathematics assessments. Literacy is generally distinguished from the traditional content domains by its emphasis on knowledge and skills required for effective participation in practical, everyday life. With the current concern over functional literacy in mind, the national assessment carried out a special study of literacy among young adults (twenty-one to twenty-five) in 1985. If the number of insightful studies that have already appeared as a result of the 1985 literacy assessment is any indication, the interest in future assessments of literacy is likely to remain high. In light of the rising level of national concern over the literacy of young adults, we recommend that NAEP continue to pursue the assessment of literacy.

One approach to assessing literacy is to include seventeen-year-olds, both in and out of school, in the school-based subject matter assessments. The special skills and knowledge tested in the literacy study could be incorporated into the subject matter assessments. Including a sample of the population of seventeen-year-olds would greatly improve the trend information on student achievement. With the full population instead of only the in-school population sampled, inferences about improvement from the thirteen-year-olds to the seventeen-year-olds could be made.

If, as we suggest (chap. 3), the assessment should be restructured so that age/grade cohorts are followed over time, then out-of-school seventeen-year-olds would be located from the information collected earlier, when they were still in school as thirteen-year-olds. We recommend that literacy also be assessed beyond high school, perhaps at age twenty-one and again four years later, at age twenty-five. This cohort design would allow the assessment to follow individuals beyond high school age, permitting the study of adult literacy in terms both of prior school background and present employment experience. There are two strong arguments that support this expansion: first, it would yield information about the type and extent of different literacy-related problems currently faced by different age groups in the adult population; second, comparison of age-group differences would enable researchers to identify the types of employment and other social experiences that are associated with various degrees of literacy across the population.

MATHEMATICS, SCIENCE, AND TECHNOLOGY

The ability to make use of higher-order thinking skills in mathematics and science is of particular importance to the future well-being of our nation. Further growth in our economy, as well as the future vitality of our way of life, will depend in large part upon continued technological and scientific progress. As a nation, then, we must assure that each generation is able not only to duplicate the accomplishments of the past but to devise creative technologies for the future. Our society must further see to it that its members are well prepared to meet the challenges of an increasingly competitive international marketplace. The founders of NAEP recognized these realities when they included the regular assessment of mathematics and science as an integral part of the initial national assessment design. Since that time NAEP has earned a well-deserved reputation for the precision of its content objectives and for the relevance of its science and mathematics assessments to educational policy.

Until very recently, however, NAEP has focused most of its attention on the assessment of achievement outcomes that are easy to measure and report statistically. Although these memory and procedural skills are important to achievement, they are actually the final products of higher-order processes that select and organize the information necessary for problem solving. Without these preliminary organizing skills, students cannot learn to match problem situations to the procedures best suited to their solution. In part as a response to national assessment information, mathematics and science educators have increasingly shifted their teaching emphasis away from the memorization of concepts and calculation routines, and toward the more demanding skills of estimation, conjecture, and problem solving. Future assessments in these two subject areas should reflect these changes, or NAEP will risk inhibiting further progress in this direction. The national assessment should move beyond the measurement of lower-order procedural skills to the assessment of how well American students are thinking and reasoning in mathematics and science.

Mathematics

Until recently, national assessments of mathematics have maintained a multiple-choice format and have measured the abilities of students to carry out explicit procedures and computations. These tasks do not require the student to point out the information most important to the goal of solving the problem or to develop strategies for using their arith-

metic skills effectively. With their advantages of economy and ease of scoring, multiple-choice tests will no doubt continue to be used in measuring mathematics achievement. Future assessments, however, should attempt to discern the problem-solving strategies and complex knowledge that students bring to more demanding tasks like story problems. NAEP assessments of mathematical achievement have already begun to probe for these skills, and we encourage further progress in this direction.

To ensure continued progress the assessment should explore new avenues of item design. One promising possibility is the use of open-ended questions which allow students to explain their reasoning while solving problems. NAEP can adapt more open-ended approaches to mathematics assessment in evaluating the higher-order skills students bring to arithmetic at each age level. The assessment should also build upon its demonstrated expertise in the development of content objectives to identify major skill areas in mathematics and to measure the ability of students to coordinate skills of various levels in solving problems.

One particularly strong argument for the development of new assessment methods in mathematics is their potential value as sources of diagnostic information for teachers and educational policymakers. The Study Group encourages NAEP to integrate new assessment approaches with studies of teaching techniques and classroom environments, in order to determine which factors are most revelant to higher-order mathematical thinking.

Science

In the past, NAEP has assessed achievement in science at irregular intervals. In view of the importance of science education to the contemporary curriculum, the Study Group recommends that achievement in science be measured on a regular cycle.

Much of what we have said about the need for assessing higher-order thinking in mathematics applies equally to science. Early national assessments of science achievement focused almost exclusively upon factual knowledge, and did not probe at all for the ability to organize and transform a body of facts into a coherent scientific account. One significant departure from this approach is the recent pilot assessment by NAEP of student abilities to assemble and organize the elements of a scientific experiment.

We welcome this new initiative into the assessment of higher-order scientific thinking and encourage continuation in this direction. In particular, we recommend that NAEP assess not only the ability to organize experiments but also the initial processes by which students sort through the information available to the experimenter. National assessment also needs to identify carefully those complex thinking skills that are most important to the various fields of scientific study. A factually rich field like biology may make significantly different demands on the thinker than a more mathematically based field like physics.

As in other achievement areas, NAEP will need to broaden its repertoire of assessment techniques if it hopes to sample higher-order scientific reasoning directly. Open-ended and free-response questions may serve as relatively economical alternatives to the multiple-choice format. These types of questions, not difficult to code, require the student to generate the correct answer, not merely to recognize it. Such assessment items would in turn allow for more reliable inferences about the thought processes contributing to the answer. One particularly promising use of a free-response format, for example, might involve the assessment of the ability to develop coherent explanations for a set of causal relationships.

We therefore recommend that the national assessment explore new methods with the evaluation of higher-order scientific thinking firmly in mind. It will cost more to develop and implement this expansion, but we see these costs as a necessary and timely investment in our nation's economic and technological future. Further, the assessment should improve its measurement of student attitudes toward the study of science.

Technology

Computers and computerized equipment are becoming common features of American life. Computer competence requires a number of general skills, including the capacity to adopt new symbol systems and put them to use. The national assessment very recently concluded its first special evaluation of computer competence as a part of its general assessment of scientific understanding. We strongly encourage the assessment to make the measurement of computer skills a continuing concern. We recommend in particular that the assessment identify and evaluate a number of general skills of central importance to the use of new technologies.

HISTORY, GEOGRAPHY, AND CIVICS

From time to time, the national assessment has investigated what children know and can do in content areas beyond the regularly assessed

core. These occasional special assessments have often occurred in response to new needs for information within the educational community, and they have contributed an interesting and valuable dimension to the overall national portrait of student achievement. For example, the assessment from time to time has examined some of the elements of the school curriculum which contribute to citizenship education, including knowledge of American history, geography, and civic affairs. The Study Group recommends that these subjects be considered part of the core area and that they be assessed on a regular basis.

As with the other regularly assessed subjects, we strongly urge that these assessments move beyond simple measurement of factual knowledge toward the assessment of more complex thinking skills. Achievement in such areas as history and civics, for example, may require not only a command of facts but also an ability to pose and analyse a well-conceived problem, often one involving value judgments of some subtlety.

SPECIAL ASSESSMENTS

Among other areas that have been the object of special assessments in the past are music, art, literature, and occupational attitudes. We urge the continuation of such efforts on an occasional basis, and we foresee that other special assessments can and should be undertaken from time to time in response to the expressed needs of the public and the educational community. Again, assessments in the arts should focus not only on historical knowledge and attitudes but also on the more complex modes of thought involved in artistic composition.

3

Structure and Design

The new responsibilities and the changes in content and coverage that we have recommended will require modifications in the technical design of the national assessment. In addition, the Study Group is suggesting other changes in structure and design which are intended to enhance its effectiveness. These include using computers for testing, improving the achievement scale scores and the usefulness of NAEP data for research and policy, following students over time, and improving the linkages of NAEP with other data-collection efforts.

NATIONAL, STATE, AND LOCAL ASSESSMENTS

The call for representative data covering all states has important implications for the structure and design of the assessment. Sample sizes and field administration will need to be greatly expanded, and the dissemination of results will have to be restructured in order to inform a much larger audience. The details of how these tasks will be accomplished are beyond the scope of this report, and will have to be addressed by the Educational Assessment Council in conjunction with the contractor. Our recommendations are confined to the content areas that should be assessed at the state and national levels, and the opportunities that should be made available to state and local jurisdictions for using the national assessment for their particular purposes.

The Study Group recommends that the new NAEP assess reading, writing, literacy, mathematics, science, technology, history, geography, and civics on a state-by-state basis. The report card will present average scores in these content areas for the nation as a whole and for each of the fifty states and the District of Columbia.

The assessment has periodically measured student performance in a number of other areas, and we have urged that this practice be maintained. We recommend that the additional subjects be measured at the national level only. This would involve a smaller sample of students than the state-by-state assessments, and would not yield individual state scores. States should be given the option of adding on to this national data base at their own expense, making possible either broader comparisons with the national sample or more refined intrastate comparisons.

Some policy issues that the assessment has addressed in the past cannot, indeed, be adequately examined within the separate states. For example, the assessment data have been used to identify trends in the performance of blacks and Hispanics, to compare the achievement of public and private school students, and to analyze the effects of different school policies on student performance. A number of states, however, have very small numbers of minority and private school students, and some school policies are decided at the state level and thus do not vary among schools within states.

Given the national importance of these and other issues, we recommend that the national sample continue to provide information that can answer questions about the achievement of policy-relevant subgroups, and about the processes that generate achievement differences. Again, states and localities should have the option of augmenting the state or national samples and of carrying out such studies on their own, at their own expense.

POPULATION AND ASSESSMENT CYCLE

For nearly twenty years, NAEP has provided trend data on educational attainment which have been unmatched in size, scope, and precision by any other national data-collection effort. Sampling and scheduling decisions, however—usually necessitated by funding shortages—have sometimes led to deficiencies and detracted from the quality of the national assessment. In what follows, the Study Group makes several recommendations aimed at overcoming these deficiencies, conscious that each will require a substantial increase in funding.

Cohort Sampling

The Study Group recommends consideration of changes in sampling procedure to allow assessment of the same individuals at four-year intervals. This change would offer many advantages over the current procedure of drawing entirely new samples for each assessment. Following the same group of students would promote far better understanding of student learning, providing a badly needed historical record of how given cohorts progress through school.

Under the new procedure, the first assessment would occur when students are nine years old or in the fourth grade. The same cohort would be assessed four and eight years later. The Study Group recommends that every two years a new nine-year-old/fourth-grade cohort be

sampled and tested in one or two subject areas. For the assessment of literacy, cohorts might be followed beyond secondary schooling and assessed at age twenty-one and again four years later at age twenty-five.

Sampling Out-of-School Seventeen-Year-Olds

The Study Group strongly recommends that out-of-school seventeen-year-olds be included in the regular assessments. In the early years of the assessment these individuals were included, but subsequent funding cuts eliminated them. The inconsistent record on this segment of the population represents a serious deficiency in assessment data. Since dropout rates change over time, the composition of the in-school student body also changes. This casts doubt on the validity of comparisons of seventeen-year-old achievement levels with those of earlier cohorts, and so disrupts the accuracy of the historical record. The importance of including this group is accentuated by the fact that individuals who drop out of school confront many problems that the assessment could usefully document.

Age/Grade Sampling

In order to preserve the historical record of the assessment, sampling of the traditional nine-, thirteen-, and seventeen-year-old groups should be preserved. The recent link of age to grade sampling should also be maintained. We recommend, however, that the assessment consider shifting the grade level samples from the current third, seventh, and eleventh grades to the fourth, eighth, and twelfth grades. This change is justified on two counts. First, the eighth and twelfth grades represent key transition points in American education, usually the end of junior and of senior high school. Second, the proposed shift would facilitate comparisons between results from the assessment and the National Educational Longitudinal Studies (see the last section of this chapter), and thus provide important checks on the validity of conclusions drawn on the basis of one or the other data set.

NEW ASSESSMENT TECHNOLOGIES

Developments in computer technology are rapidly changing the cost and quality of data collection. The national assessment should consider making substantial investments in this new technology with the idea

of reducing the long-term cost and increasing the quality of the assessment data.

In particular, replacing paper-and-pencil with computerized testing, which is now feasible, would offer a number of benefits. For one, it would facilitate the study of students' thinking processes. The actual process of thinking that a student follows when faced with a problem is difficult or impossible to identify on the basis of responses to the current multiple-choice, paper-and-pencil tests. Advances in computer technology and in our understanding of problem-solving behavior now make it possible to identify where students "go right" or "go wrong" with a task, and can thus provide valuable clues for improving instruction.

New technologies also allow for more efficient use of time. Accuracy equal to paper-and-pencil instruments can be obtained with greater precision in significantly less testing time and with many fewer items. This in turn would allow the assessment to extend the current range of content covered.

Finally, computerized testing could shorten the time between the collection of data and the dissemination of results. With the expansion to state-level collection and reporting, methods of expediting the dissemination of information will become particularly important.

In light of these anticipated benefits, the Study Group recommends that resources should be devoted to the development and implementation of a computer-based national assessment system. This system should be fully operational within the next decade.

A NEW TEST DESIGN

In contrast to many achievement testing programs, NAEP does not administer the same set of items to all students sampled within an age-grade group. Students participating in the assessment are instead given different samples from a large pool of items within a content area. This procedure makes it possible to assess a broader range of skills than could otherwise be done in the testing time currently alloted. The drawback of the method is that individual students cannot be directly compared, since they have responded to different sets of items. While comparisons can be carried out among groups of students, analyses of variability in student achievement cannot be conducted. This limitation has greatly reduced the use of NAEP data by educational researchers, resulting in an important loss of information on factors that affect student learning.

In order to increase the usefulness of national assessment data, proposals for improving the form design have been advanced. Among

these the "duplex design" proposed by Bock (1986) would improve the reliability of the scales without significantly increasing individual student testing time. But if, on the other hand, student testing time were increased, the perceived needs could be met by that means. In this regard, the Study Group concludes that testing time could indeed be substantially expanded without endangering school cooperation. Either approach would result in speedier access to and wider use of the assessment data. Our recommendation is for an increase in testing time.

STUDENT BACKGROUND AND SCHOOL VARIABLES

Historically, NAEP has collected some information on characteristics of respondents' communities, including the region of the country in which such community is located, its size, and socioeconomic status. NAEP has in addition measured a few student background variables, such as race and ethnicity, age, sex, and parents' educational attainments. This set was expanded in the 1984 assessments to measures of students' homework and television–viewing habits, and the availability of reading materials in their homes. NAEP has begun to gather school-level information from principals and teachers, focusing particularly upon course offerings, teaching methods, and staff training and experience. Plans are now underway to collect more information on instructional practices, including ability grouping, textbooks used, and content coverage.

NAEP has sought to redress the criticism that it collects insufficient background information by enlarging the number of variables measured. The Study Group is now concerned, however, that the collection of greater volumes of background and contextual variables might entail the risk of lowering the rate of participation among schools and students. Of even greater concern are the questions of how relevant the present background and contextual variables are to matters of educational research and policy, and how readily this information can be translated into meaningful action toward the improvement of education.

The collection of school variables promises to improve substantially the value of the assessment as a policy research tool, and the Study Group recommends that this information be gathered on a regular basis. The Study Group strongly recommends, however, that school variables collected by the national assessment be demonstrated by previous, smaller-scale studies to have significant effects on student achievement. Otherwise, there are real dangers of making an already large and complex project even more formidable to data analysts, of escalating collection and processing costs, and of reducing participation of schools owing to the added response burden. It is crucial then that the assess-

ment should not try to play an exploratory role with respect to variables affecting achievement, and the new Educational Assessment Council should see that it does not.

LINKING WITH OTHER DATA COLLECTIONS

As the national assessment has developed, many of its functions have begun to overlap with other federal data-collection efforts. Some of the recommendations made in this report, particularly the move to state-level sampling, would increase the overlap. In view of interest at the newly reorganized Department of Education in coordinating its data-collection projects, the Study Group considered a number of proposals for linking the national assessment with two other large projects, the Elementary/Secondary Information Data System (ESIDS) and the National Educational Longitudinal Studies (NELS).

The ESIDS currently collects a wide array of data on state and local school systems, including information on staff characteristics, finance, enrollments, and school facilities (see Appendix E of this report). As the national assessment has begun to gather information on schools and school districts, and as the move to include state-representative samples appears imminent, the feasibility of linking the two studies has increased.

The national assessment and NELS also have a number of common objectives. Both collect national representative data on student background and outcomes, and on characteristics of schools and students' experiences within them. Both are concerned with assessing and explaining changes in student academic achievement. The two studies differ, however, in terms of their designs. In contrast to the assessment, NELS samples single-grade cohorts and follows the selected students as they progress through school and their post-secondary careers. The followups are usually carried out every two years, and studies have been conducted over a ten-year period. NELS has initiated three such studies to date: the National Longitudinal Study of the 1972 senior cohort, the High School and Beyond Study of the 1980 sophomore and senior cohorts, and the NELS 88 study of the 1988 eighth-grade cohort.

The Study Group recommends that NAEP pursue linkages with these two important projects. The comparability of NAEP and NELS data would be improved if the assessment samples fourth-, eighth-, and twelfth-graders. The high–quality achievement data gathered by NAEP could be linked with the extensive information on schools collected by ESIDS if these projects aligned their school samples. The Educational Assessment Council should routinely confer with NELS and ESIDS staff concerning sampling, questionnaire and test development, and survey

administration. Greater linkage of the three projects promises to make the data collected more useful for analyzing relationships among achievement, school policy, classroom practices, and other background information.

Note: After this report, Linda Darling-Hammond, a member of the Study Group, requested that the following statement of her reservations be included:

"The Study Group has made sweeping recommendations for reshaping and, hopefully, strengthening the National Assessment of Educational Progress. I endorse the report's intentions and its recommendations to improve the quality of NAEP. However, I remain concerned that the effort to make NAEP data useful for a greater range of purposes will undermine the assessment's capacity to perform its basic mission effectively.

"Dilution of resources and distortion of purposes can result from extensive use of NAEP for district or school-building comparisons, or from efforts to link NAEP to other assessments or data collection efforts. These avenues should, in my view, be pursued with great caution. There is a delicate balance between developing a first-rate assessment of what the nation's students know and can do and attempting to negotiate a multi-purpose testing and data collection effort that may satisfy many objectives superficially but none of them well. Top priority should be given to improving the quality of the assessment and to safeguarding its role as the nation's report card.

"As the study group's recommendations are considered, I hope the temptation to use NAEP as an all-purpose testing and data collection tool will be resisted, and that changes in the assessment will be designed to improve its ability to serve its important functions with integrity."

4

Governance

Federal law now provides that the contractor for the national assessment appoint a committee to establish policy for that particular assessment. Since we foresee a much weightier role for such a policy body, and a need to assure continuity of policy formation when contractors change, we recommend revising the law to establish an independent Educational Assessment Council, to be appointed by the Secretary of Education from nominees selected by a broadly based nominating committee. In addition to setting policy, the Educational Assessment Council (EAC) would design each periodic assessment.

RESPONSIBILITIES OF THE EDUCATIONAL ASSESSMENT COUNCIL

The chief responsibility of the new council would be to shape each assessment, selecting the content areas to be tested, defining conceptually the ground to be covered in each area, setting test specifications, and identifying feasible achievement goals for each of the age and grade levels to be tested. Closely linked with this activity would be the EAC's responsibility to set policy—on such matters as maintaining the continuity over time of the assessment's data banks, setting standards and procedures for the use of test data, and arranging for periodic review of the whole assessment effort by outside bodies (such as the present Study Group). In addition the council would explain, interpret, and represent the national assessment to educators, government officials, and the general public, assisting them in their use of its invaluable resources.

In carrying out these tasks, the council, supervising a permanent staff of perhaps ten professionals, would work with state and local curriculum specialists and test directors, as well as other experts in each field of study. Advisors would be constituted in specialized panels, addressing such issues as selection of educational objectives, construction of assessment exercises, sampling procedures, reporting and interpreting student performance, and the potentials and limitations of educational assessments. As the national assessment expands its responsibilities to state and local educational authorities, advisory panels should also provide guidelines for state and local linkages with the national data, including standards for fair and constructive comparisons, and performance expectations. Each panel would be charged with providing the council with the technical knowledge required for policy decisions.

The federal government would award a grant to an independent organization chartered by law to house the Educational Assessment Council.

RESPONSIBILITIES OF THE TEST CONTRACTOR

The federal government will select, fund, and monitor a contractor to conduct the national assessment. The contractor would be responsible for all item development, test and questionnaire administration, processing, and analysis, and would release findings to the states and the federal government, disseminate them, and provide technical assistance to states and local districts who wished to link their assessments with national data. The test contractor would also respond to state and local requests for additional assessment data collections to be paid for by states and localities. All of these responsibilities would be conducted in compliance with the general policies, content areas, and objectives prescribed by the Educational Assessment Council. The council would also specify, in connection with each content area or skill, whether it was to be tested on a state-by-state basis or nationally.

ESTABLISHING THE COUNCIL

The Educational Assessment Council, as we envision it, would be a highly responsible and prestigious group of men and women, themselves expert in their fields and prepared to requisition and use the expertise of others. The members, appointed by the Secretary of Education, should serve staggered terms to assure continuity, terms of perhaps five years, after the initial appointments. They might well have had experience in such positions as the following (though not necessarily hold that post currently):

- Governor of a state
- Chief school officer for a state
- State legislator
- Local superintendent
- Member of a state board of education
- Member of a local board of education

- Principal
- Classroom teacher
- Representative of business and industry
- Curriculum planner or supervisor
- Testing and measurement expert
- Private school respresentative
- Education researcher

Ex officio members of the council will include a member of the National Advisory Council on Educational Research and Improvement and the Assistant Secretary of Education for Educational Research and Improvement.

To provide the Secretary of Education with lists of qualified candidates for membership on the Educational Assessment Council, a permanent standing committee would be established by federal statute. This nominating committee should represent a range of professional perspectives in both government and education. Our recommendation is that the committee be composed of the current incumbent of each of the following posts:

- Chairmen of the House and Senate education committees
- Chairman of the National Governors Association
- President of the Council of Chief State School Officers
- Chairman of the National Council of State Legislators
- President of the National Academy of Education
- President of the National Academy of Sciences

The nominating committee would solicit names from their colleagues, from other organizations interested in education, and from the public at large, and should forward to the Secretary of Education at least three nominations for each position to be filled.

ADVANTAGES OF THE NEW STRUCTURE

The Education Assessment Council we propose would differ from the present Assessment Policy Committee in several important respects. First, we believe that it is essential to separate the Educational Assessment Council from the test contractor. This would establish a set of checks and balances among the three entities involved in the assessment: the council, which sets testing policy and test specifications, the test contractor, who develops and administers the actual tests, and the federal government, which provides funding and awards the contract. We believe that negotiations among these three groups will strengthen the

decision-making process by reflecting an array of education, measurement, and policy perspectives.

Another important advantage is the ability of the new Educational Assessment Council to provide continuity beyond the duration of any given contract. Continuity is especially critical in view of the broader functions that we envision as within the council's charge. In addition to specifying the dimensions of the actual assessment, the council will be responsible for policy statements on such matters as the limitations and fair use of assessment data. It will help interpret the nation's report card for policymakers, educators, researchers, and the general public, making clear the processes through which these findings are derived. Continuity in assessment policy and clarity in explaining it become particularly critical as other long-range data-collection efforts are more closely linked with the national assessment.

It will of course take time for the new structure envisioned in this report to become fully operational. Only modest changes may be effected in time for the 1988 NAEP assessment. But if we assume successful discussions with Congress, the new structure could be in place and functioning in time for the 1990 assessment.

5

Costs

The new assessment design envisioned by the Study Group will increase the federal investment to about $26 million a year. The largest item, collecting information on a state-by-state basis, will demand about $13.5 million of this. Adding transcript data on twelfth graders, out-of-school seventeen-year-olds, and parent data for nine-year-olds or fourth graders will come to about $3.25 million, and development of test objectives with the assistance of state and local experts, expanding item development and data banks, and research on technical development in assessment, about $2.6 million. Analysis, dissemination and reporting, technical assistance to users, and general management will require some $4 million. The Educational Assessment Council, with its professional staff, will demand about $2.5 million. Thus the federal outlay for the new design will total a little under $26 million a year, an investment that we believe will provide the nation with immensely rich returns in the form of improvement of education. Our estimates are summarized in Table 1 and discussed in greater detail below.

It may be noted that none of these projected cost estimates include the costs of state "add-ons," since these are paid for by the states requesting them. The purpose of an add-on is not for state-to-state comparisons, since these are obtained from the main NAEP assessment data, but rather for more refined analysis within a particular state. State-administered multiple-choice assessments could be conducted for approximately $150,000 per grade per content area.

TABLE 1

The New Report Card: Estimated Annual Costs

Component	Cost
1. Field data collection	$13,471,000
2. Transcript data	750,000
3. Data on out-of-school 17-year olds and older groups	1,500,000
4. Data on parents	1,000,000
5. Objectives development	600,000
6. Item development	1,000,000
7. Assessment technologies development	1,000,000
8. Analysis	1,925,000
9. Dissemination and reporting	995,000
10. Supplying technical assistance	550,000
11. Management	550,000
12. Educational Assessment Council	2,500,000
Total	$25,791,000

1. Field Data Collection

Assumed here are a sample of 4,500 students per grade per state, distributed among public and private schools, with an average cluster size of fifty students per school, and three hours of data-collection time per student. It is also assumed that most (but not all) data will be collected state by state. The costs as estimated break down as follows:

Student data (688,600 students at $20/student)	$13,770,000
School data (13,770 schools at $600/school)	8,262,000
District data (12,000 districts at $400/district)	4,800,00
State data (51 "states" at $200/"state")	10,200
Fixed costs	100,000
Total for two-year cycle	$26,942,200

Or, as given above, approximately $13,471,000 per year.

2. Transcript Data

During every two-year cycle, information on courses taken would be collected from the transcripts of students in grade 12. These data would be collected for a national sample of approximately 30,000 to 40,000 students.

Cost for a two-year cycle is about $1,500,000, or $750,000 per year.

3. Out-of-School Seventeen-Year-Olds and Older Groups

Since data for out-of-school seventeen-year-olds are very difficult and expensive to collect, such data should be gathered only every four years. For the national sample of 1,500 out-of-school seventeen-year-olds NAEP content areas will be assessed. For other age groups data collection will be limited to literacy assessment.

Cost for a four-year cycle is about $6,000,000, or $1,500,000 per year.

4. Parent Data

Since many nine-year-olds and fourth graders are unable to supply reliable background information, it is desirable to have comparable data reported by parents. A national sample of about 30,000 parents of the youngest age/grade group would be selected and asked to respond to a questionnaire eliciting background data.

Cost for 30,000 parents at $65 each is $1,950,000, or about $1,000,000 per year.

5. Objectives Development

State and local curricular experts would be brought together to develop objectives for the different content areas to be assessed. It is assumed that, on average, four areas will be assessed in each two-year cycle. First, the Education Assessment Council staff would analyze objectives and curricula submitted by the states. A national meeting of representatives from the states would be held to begin the review process, with four regional meetings to follow that would include state and local representatives. A second national meeting would be held to summarize the process and produce objectives for the assessment. This procedure would be carried out for each content area.

For each content area six meetings at $50,000 would cost $300,000. So work on four content areas would come to $1,200,000 over the two-year cycle, or $600,000 per year.

6. Item Development

With the proposed assessment strategy, especially the provisions for state involvement, considerably more effort will have to go into new item development. It is anticipated that more users of NAEP items will increase the demand for new and higher-quality items.

Cost would be approximately $2,000,000 for a two-year cycle, or $1,000,000 per year.

7. Assessment Technologies Development

The current grant includes no funds for studying how the assessments are carried out (evaluation and development funding). Three developmental tasks need to carried out on an ongoing basis:

1. Examination of NAEP item banks and sampling for state and local uses
2. Research into the effects of assessment design on state and local uses of assessment items
3. Research on such topics as the utility of duplex design, the possible effects of the timing of assessments, and the like

Cost would be $2,000,000 for each two-year cycle, or $1,000,000 a year.

8. Analysis

Over a two-year cycle the data will need to be analysed to produce at least eight to ten detailed national reports and summary reports for each of the states. Analysis will include processing, editing, scaling, and tabulation.

Cost would be $3,850,000 for a two-year cycle, or $1,925,000 a year.

9. Dissemination and Reporting

It is expected that at least ten national reports with press conferences, news releases, and follow-up activities would occur during a two-year cycle. Each such activity would cost about $100,000, and providing and disseminating data tapes for public use would add perhaps $12,000, for a total of $1,120,000.

In addition, state reports, at a cost of about $10,000 each, would be required with tapes also provided to the states, for a total cost of $520,000 over a two-year period.

In order to encourage school participation, summary school reports would be provided for all 13,770 schools at around $25 per school, or $350,000 for every two-year cycle.

So the cost for national, state, and local dissemination activities over a two-year cycle would be about $1,990,000, or $995,000 a year.

10. Technical Assistance

Assistance will be offered to national, state, and local users of the assessment data. This will require the services of three full-time staff members with travel, telephone, and clerical assistance.

Cost would be $1,100,000 for a two-year cycle, or $550,000 per year.

11. Management

Management of a project of this magnitude would take the full-time attention of approximately four senior staff people, with clerical support.

Cost would be $1,100,000 for a two-year cycle, or $550,000 per year.

12. Educational Assessment Council

The Educational Assessment Council, proposed as the policy-making body for educational assessment activities in the United States, will need to meet about four times a year. In addition to these meetings of the whole council, meetings of subcommittees averaging three members each would have to take place about ten times a year.

Cost would be about $2,000,000 every two years, or $1,000,000 each year.

The Educational Assessment Council will require a full-time staff of ten professionals, with secretarial support. It is suggested that the staff include subject-area specialists in reading, writing, mathematics, science, and other content areas as specified by the council. The staff should also include policy and research specialists and an assessment expert. This staff would provide support to the council and carry out activities commissioned by it.

Staff costs would be about $3,000,000 for every two-year cycle, or $1,500,000 each year.

So the total costs for work of the Educational Assessment Council, including staff costs, would come to $5,000,000 for a two-year cycle, or $2,500,000 per year.

Part 2

Commentary by the National Academy of Education

THE REVIEW COMMITTEE OF THE NATIONAL ACADEMY OF EDUCATION

Robert Glaser
University of Pittsburgh
Chairman

Ann L. Brown
University of Illinois

Edmund W. Gordon
Yale University

Harold Howe II
Harvard University

Albert Shanker
American Federation of Teachers

Lee J. Cronbach
Stanford University
Consulting Advisor

Report prepared by
Anthony S. Bryk
University of Chicago

ACKNOWLEDGMENTS

This commentary on the Alexander-James Study Group's report was made possible by funds provided by the Carnegie Corporation of New York, the Andrew Mellon Foundation, and the Rockefeller Foundation. We are grateful for their support of the Academy's review. We also wish to extend our own note of gratitude to those foundations, previously acknowledged in the Alexander-James report, who contributed to the support of the overall project. The committee notes its personal indebtedness to H. Thomas James and his staff. Mr. James kept us regularly informed about the panel's deliberations, provided access to the extensive materials developed by the panel, and generally facilitated our review. His contributions were characterized by generosity of spirit and goodwill, without which this coordinated effort could not have been accomplished. Lastly, we owe a vote of thanks to Gail Keeley, executive director of the National Academy of Education, who provided staff support for the committee. Our work proceeded under a very tight time frame, and Gail assured that we were all in the right place, at the right time.

Commentary

Schools play a critical role in our society. In forming that educated citizenry necessary for the sustenance and growth of our democratic charter, schools provide the foundation on which our nation rests. This observation is of longstanding historical merit, although it is now largely ignored in public discussions about schools. Current discourse is dominated by a narrow utilitarian conception of purpose that sees the aims of schooling principally as advancing the ability to compete in the world marketplace. We must be always mindful, however, of the broader purposes that schools serve and the real threat that weak schools pose. Quite simply, no democracy can long survive that does not assure its children an education that allows them to take full advantage of the freedom and opportunities which a democracy offers.

AN APPRAISAL OF NAEP: PURPOSES AND DIRECTION

It is for this reason that the lack of support and, in recent years, fiscal starvation suffered by the National Assessment of Educational Progress appears so troublesome to us. The current budget of the national assessment is considerably less than it was in the 1960s—in constant dollar terms, a small fraction of the support originally provided for it. The current state of NAEP is indicative of pervasive and longstanding problems in the collection of educational statistics. A recent report by the National Academy of Sciences shows that during the past ten years, a period of rising demand for more and better information about education, federal staffing and budgets for assessment activities have declined significantly. It is puzzling that the federal government routinely supports the collection and dissemination of statistics on the condition of our economy, our health system, and even the incidence of crime, but places so little priority on the condition of our schools.[1]

A regular, thorough examination of schooling is one of the federal government's most important tasks in education. This sense of the nation's need for good information about its schools was established by act of Congress in 1867. It remains important today.

Thus, the Academy Committee is encouraged by the renewed interest in the National Assessment of Educational Progress that has been demonstrated in the work of the Alexander-James panel. Among its members were a broad cross-section of American educators, businessmen, and elected officials all sharing a common purpose—to improve our nation's schools through a systematic assessment of educational progress. They drew broadly on outside consultants to bring both technical and subject-matter expertise to their deliberations. The panel undertook an ambitious and timely task, and we applaud them for the energy that they brought to this effort. We also commend the vision of the Secretary of Education in initiating this activity to improve NAEP.

The National Assessment of Educational Progress, by focusing attention on our schools and by contributing new information, can be a positive force for improving the quality of our children's education. In this regard we share a common purpose with the Alexander-James Study Group, and we offer the comments below in a spirit of constructive discourse.

NAEP AS A MODEL OF WHAT STUDENTS SHOULD KNOW AND HOW IT SHOULD BE ASSESSED

The nation has a right to know what students achieve, what schools are doing, and what more should be done. Information on student progress, wisely interpreted, is of obvious value to the public, to educators, and to policymakers at all levels of government. A well-designed NAEP can help to focus instruction and educational practice on meaningful learning activities. NAEP should enable us to ensure that students are learning something of value and help us to assess adequately the fruits of that learning.

We stand at an exciting point in the development of techniques in measurement and assessment. The widespread availability of statistical computing has changed the very countenance of quantitative social and behavioral science. We now routinely perform analyses that were literally impossible less than twenty years ago. These developments have been particularly significant in the theory of test construction. The recent NAEP report on literacy, *Profiles of America's Young Adults*, offers positive testimony in this regard. Unlike conventional standardized tests, which rely almost exclusively on a multiple-choice format, respondents in the liter-

These comments were prepared in response to a draft of the Alexander-James report. The Introduction and summary preceding the full report was written just prior to publication, and consequently was not available to the Academy during its review period. We encourage those concerned about the future of NAEP to examine the full report of the panel.

1. *Creating a Center for Educational Statistics: A Time for Action.* Washington, D.C.: National Academy Press, 1986.

acy assessment were asked to demonstrate competencies in a variety of ways. The tasks examined called on important skills required to function in a modern technological society. Through use of recent developments in item–response theory it was possible to map the diverse items employed in these tests into coherent scales, and further, to equate these results with those of more conventional reading tests administered to school-based samples.

Such developments in techniques of measurement and in the analysis of human competence and cognitive skills advance the state of assessment. NAEP should continue to take a position of leadership in establishing standards that both define the knowledge and skills to be assessed and specify how these assessments are to be undertaken. The background papers generated by the Study Group suggested many additional opportunities meriting further consideration. Some are substantive, such as an increased emphasis on assessment of higher–order thinking skills across all content areas; others are methodological, such as developing use of microcomputers in testing situations. In this time of heightened interest in and use of assessment, we should expect NAEP not only to employ the best available techniques but also to contribute actively to advancing the state of the art. Toward these ends sufficient resources must be provided to support adequately the necessary research and development efforts. Not only will the nation's report card benefit, but the standards set by NAEP can have pervasive effects on state and local assessment systems as well.

NAEP AS A CATALYST FOR SCHOOL IMPROVEMENT

The time seems appropriate to expand the purposes of the national assessment beyond simply reporting on the status of learning in the nation. NAEP results should be more than mere causes for alarm or reasons for complacency. These reports should help to inform discussions about the improvement of schools. While we support the general tenor of the Alexander-James report in this regard, it is important to recognize that this is a new task for NAEP and one whose pursuit requires careful deliberation.

Several of the recommendations of the Alexander-James report are of obvious merit in this regard and deserve strong endorsement. The expansion of NAEP to include regular assessments in science, history, civics, and geography in addition to the basic skills of reading, writing, and mathematics is clearly warranted. The change in design to facilitate the collection of data and reporting of results for individual states and

localities is another important development. The commitment to continuity of future data collection with that of the past is essential to ensure the validity of comparisons over time.

What remains less clear in the panel report is how NAEP data will actually link to school improvement efforts. Although NAEP can tell us a great deal about "how our schools are doing," it provides only limited and mostly indirect evidence about the factors contributing to these successes and failures. It is natural to suggest that NAEP data collection be expanded so as to shed more light on these causal linkages. Unfortunately, few such questions are well suited for examination within the current NAEP design. For example, the Alexander-James report suggests, with good reason, that teacher characteristics and classroom emphases are major causes of educational outcomes. In the present NAEP design, however, pupil progress from one assessment to the next reflects the cumulative effect of several teachers and varied classroom experiences over a four-year period. To relate pupil performances to the average characteristics of the teachers within a school greatly attenuates the estimates of the true effects. Students experience only a few of these teachers, and both pupils and teachers move in and out of schools. Serious research on these questions requires a design that is faithful to the temporal and structural features that frame the learning activity. Academic learning is a dynamic process that principally occurs within a classroom during an academic year as teachers and students interact over specific curricular materials. These features imply a need for a more finely detailed analysis than is possible within the current NAEP. In fact, this basic research is probably better pursued as a separate enterprise within the larger educational research community than as a small add-on to a large federal effort whose principal purpose is quite different.

In general, efforts at school improvement confront complex issues that cut across diverse areas of expertise. As outlined above, some of these are technical problems encountered in attempts to draw valid inferences about the causes of school success (and failure) from survey data such as those of NAEP. Some are also important organizational issues. As we later discuss in more detail, past experiences clearly indicate that when the results of testing programs are closely linked to future school activities, they may have severe negative consequences on the actual operations of the schools. More generally, we have learned a good deal over the past twenty years from attempts to use external knowledge to promote reform in schools. Much of this is evident in the working papers prepared for the NAEP Study Group. What is needed now is more time to reflect on this accumulated wisdom. Some leads can best be pursued by

the advisory panels envisioned under the new NAEP governance structure. For others it may be more appropriate to establish interim study groups that include NAEP representatives but are essentially outside the NAEP structure.

NECESSARY CONDITIONS FOR FUTURE SUCCESS

While the Study Group report offers a ringing endorsement of NAEP and promises a better future, several important conditions must exist if these visions are ever to be realized.

First, any new governance arrangement established for NAEP must safely secure the assessment from inappropriate political interests. A public system of checks and balances can help to assure the integrity of this enterprise. While some agencies have a long history of independent operation, the government's record in other areas offers less reason for optimism.

Second, it is essential that the internal organization of NAEP be structured to ensure full involvement of diverse perspectives. Education in America is a highly pluralistic enterprise. This diversity of views must be represented in choosing the skills and subject matter to be tested, in developing the test items, and in the reporting of results. Reporting is especially important as NAEP moves from simple descriptive assessment of "how we are doing" to the more analytic—and by definition, interpretive—findings on which school improvement efforts might draw.

Third, the ability to link future data with past results is essential and must be preserved. The most important feature of NAEP is its ability to compare results over time. Although the content and techniques of assessment will evolve with social and scientific progress, a consistent baseline must be maintained in order to index change.

Fourth, sufficient resources must be provided not only to maintain the assessment, but also to support related research and development activities. The current NAEP draws on a broad base of past educational research and technical development. NAEP cannot achieve the new goals set out by the Study Group without a sustained commitment to support these related efforts.

We elaborate on each of these points later in these comments.

LIMITATIONS AND MISUSES OF STANDARDIZED ASSESSMENTS

We must be ever mindful of what NAEP is and is not. We are concerned that the results of future assessments may begin to exercise an influence on our schools that exceeds their scope and true merit.

Even the best of current efforts within NAEP only provide a view of children's command of basic academic knowledge and skills in mathematics, reading, and writing. While these competencies are important prerequisites for living in our modern world and fundamental to general and continuing education, they represent only a portion of the goals of elementary and secondary schooling. There are major curriculum areas, such as the humanities, that have never been addressed by NAEP. And then there are the aesthetic and moral aims of education that remain beyond the purview of current assessment techniques.

The Academy Committee is concerned lest the narrowness of NAEP may have a distorted impact on our schools. When test results become the arbiter of future choices, a subtle shift occurs in which fallible and partial indicators of academic achievement are transformed into major goals of schooling. This possibility was recognized by a nineteenth-century British school inspector who had observed, firsthand, the negative effects of linking teacher salaries to pupil examination results:

> Whenever the outward standard of reality (examination results) has established itself at the expense of the inward, the ease with which worth (or what passes for such) can be measured is ever tending to become in itself the chief, if not sole, measure of worth. And in proportion as we tend to value the results of education for their measureableness, so we tend to undervalue and at last ignore those results which are too intrinsically valuable to be measured.[2]

At root here is a fundamental dilemma. Those personal qualities that we hold dear—resilience and courage in the face of stress, a sense of craft in our work, a commitment to justice and caring in our social relationships, a dedication to advancing the public good in our communal life—are exceedingly difficult to assess. And so, unfortunately, we are apt to measure what we can, and eventually come to value what is measured over what is left unmeasured. The shift is subtle, and occurs

2. Quoted from Walt Haney and George Madaus, "The Effects of Standardized Testing and the Future of the National Assessment of Educational Progress" (1986), a working paper commissioned by the Alexander-James Study Group.

gradually. It first invades our language and then slowly begins to dominate our thinking. It is all around us, and we too are a part of it. In neither academic nor popular discourse about schools does one find nowadays much reference to the important human qualities noted above. The language of academic achievement tests has become the primary rhetoric of schooling.

NAEP needs to expand its areas of testing to remind the nation that schooling involves much more than just basic competencies in reading, writing, and mathematics. The Alexander-James report recommends that science, history, civics, and geography be added to the core assessment. The report also suggests possible future assessments in social studies and the arts. We would like to endorse the humanities. Their importance has been neglected for too long. Here, more than anywhere else in the curriculum, is where those fundamental human qualities mentioned above are the subject matter of instruction. We repeat that what is assessed tends to become what the community values. Thus, it seems critical that the assessment direct attention toward the fullness of the human experience.

In a related vein, we are concerned that standards of minimum competence are increasingly being given the stature of ultimate and adequate goals, and as a result are exerting a strong downward press on the very educational institutions we seek to improve. We must reaffirm our loftiest ambitions for schools: they should encourage all individuals to strive toward attaining their very best. To have a productive personal and social life in an increasingly complex and technical society requires greater intellectual and personal development than ever. From this point of view literacy can only be defined as that combination of abilities and personal dispositions that enables a lifelong learner. We should settle for nothing less.

Let us now turn to some of the Academy Committee's more specific reactions to the Study Group report.

ENABLING COMPETENCIES AS THE CORE OF NATIONAL ASSESSMENT

For too long, standardized assessments in reading, writing, mathematics, and science have focused strongly on the ability to recall basic facts. These subject areas, however, also consist of important skills and dispositions that are essential to future learning. Once mastered, these become the enabling competencies for active citizenship in a technological society. A strength of the Alexander-James report is its recognition of

the importance of these enabling competencies and the need to redirect future assessments toward such educational goals.

The purpose of students learning to read is to enable them to learn from reading. In this sense mastery of reading creates a new tool for learning. Texts become sources of information that in later school years increasingly replace the teacher. Individuals who cannot decode written texts with facility have great difficulty using printed information to expand their knowledge. For too many school children and adults, reading English is laborious and discouraging. In the end many refuse to read at all. Reading as an enabling skill should be sufficiently automatic that the student can concentrate on the meaning of the text. At the most basic level, students need to be able to comprehend newspapers and magazines so that they can be informed adequately on issues of public concern and responsibility. They need to be able to follow instructions on how appropriately to use equipment, communication devices, and consumer products. In adulthood, they need to be able to read to their children and assist them with their homework. This is the basic level of competence to which schools should be held accountable and which NAEP should measure.

Similarly, writing is an essential enabling skill. It provides a way of organizing as well as communicating thought. Those who cannot express thoughts clearly are at a great disadvantage in both work and school. Functional competence certainly requires mastery of the mechanics of spelling and grammar, but it involves more—the ability to use writing to help clarify ideas and to build persuasive arguments.

Much the same can also be said about the sciences. The ability to think critically about our social and physical world—constantly to ask the questions, "What's really going on here, and why do I think that?"—enables the learner to acquire new knowledge independently. Here too enabling competencies require specialized knowledge, but also go beyond it. The very essence of the idea of enabling competencies is that they provide entry into new ways of knowing. Seen in this light, the core subject areas afford opportunities for advancing critical thinking and mental experimentation, as well as organizing and interpreting information.

HIGHER–ORDER THINKING SKILLS

We endorse the emphasis in the Alexander-James report on the assessment of higher–order thinking skills. Developments in this area provide NAEP with an opportunity to identify important skills and conse-

quently to influence future developments in curriculum. This represents the best in testing practice—to measure important objectives of schooling and thereby contribute to the advancement of those objectives. Careful work, however, will be needed in specifying these competencies and developing appropriate test situations.

It is all too easy to think of higher–order skills as involving only difficult subject matter as, for example, learning calculus. Yet one can memorize the formulas for derivatives just as easily as those for computing areas of various geometric shapes, while remaining equally confused about the overall goals of both activities. All subjects have a basic knowledge component that can be taught through drill and practice. This basic knowledge, while prerequisite to competence, is also distinct from the intellectual skills of gathering relevant information, evaluating evidence, weighing alternative courses of action, and articulating reasoned arguments. Students need to understand what they know and to reason on the basis of their knowledge, rather than just memorize and regurgitate facts on command. This concern is equally valid for all subject areas. It would be shortsighted to emphasize higher–order skills in mathematics and science while neglecting them in the humanities and social studies.

We wish to emphasize that in our view the assessment of higher–order thinking skills is best carried out within the traditional subject-matter areas rather than as a separate enterprise. Although it is possible to conceive of instructional and assessment activities directed toward principles of abstract logic and reasoning, we prefer the alternative of teaching each content area as a *form* of higher–order thinking. To remove the assessment of these skills from the context of their subject matter seems likely to trivialize them and over time may lead to their neglect in the instruction of the core subjects. Indeed, we find it hard to imagine good teaching and assessment without an emphasis on the ability to reason effectively within each domain.

SAFEGUARDING PLURALISTIC APPROACHES

By identifying and assessing functional competence, NAEP directs the nation's attention toward important educational goals. As the assessment continues to develop, however, there is the danger that it will gradually evolve into a blueprint for a national curriculum. At this point we must assert the importance of instructional pluralism. Although schools should be held accountable for their efforts, school staff must retain the ability to organize instruction in ways that they deem appropriate. Our

educational system must remain open to new approaches that promise to improve learning.

The emphasis on demonstrating functional competencies is one of the strengths of NAEP. This was nicely reflected in both the choice of exercises and testing formats in the recent literacy assessment. Respondents in this assessment were asked to demonstrate such skills as writing a short letter to correct a billing error, entering transactions in a check ledger and balancing it, and orally interpreting a lengthy feature story from a newspaper. The testing formats included reading a bus schedule, interpreting an appliance warranty, following directions from a map, and articulating information gained from graphical presentations. Each of these situations calls for skills clearly relevant to basic functioning in our modern society.

We support NAEP's efforts to develop a more diverse set of exercises that will allow students to demonstrate better their real competencies. But if the item-development task should shift toward mapping subordinate elements to be mastered on the way to developing these functional competencies, we would be troubled. The latter constitutes a blueprint for a curriculum, and presumes known answers, valid for all American communities, to questions about how children learn best. However, there is important basic research still to be pursued here to provide the requisite knowledge base. Further, it is far from clear that a single pathway to learning can ever be appropriate for all youngsters. In fact practical reason dictates that learning approaches should be tailored to varied populations and contexts.

Thus, given NAEP's position of national leadership and the far-reaching effects that its testing activities may have, NAEP test exercises should steer clear of any assumptions about particular pathways to learning. NAEP should avoid constructing item banks that envision subject-matter domains as hierarchies of skills and prerequisite subskills and facts. Although it is possible to fit such hierarchical models to test data with modern item-analysis techniques, the results of these analyses reflect both the varied ways in which children learn and the dominant modes in which instruction happens to be organized currently. Such analyses do not provide unambiguous evidence about the effectiveness of existing instruction or suggest how instruction might be better organized in the future.

On a more practical level, we should not underestimate the effort required to develop large item banks. There is a danger that in undertaking this work the time and energy of key staff will be diverted from cen-

tral purposes. Simply implementing the panel's core recommendations will involve a major expansion of effort and will be very demanding of time and talent.

A STRONG AND INDEPENDENT EDUCATIONAL ASSESSMENT COUNCIL

The Educational Assessment Council (EAC) is the central organizational entity that will direct the future course of NAEP, and it is critical that it be wisely established. In particular, the enabling legislation for the new NAEP should clearly assure its independence.

The actual relationship of the Secretary of Education to this new council remains somewhat ambiguous in the Alexander-James report. It is not clear whether the secretary would be constrained to frame the testing contract according to the "policies and specifications" set by the EAC or whether the secretary could regard these policies and specifications as merely advisory and ignore them. Since this issue remains unclear, it is only prudent to assume that the latter possibility exists. If so, the entire endeavor is left open to possible inappropriate intrusion.

We recognize that a governmental agency cannot allow an independent organization such as the EAC to dictate to the secretary the extent of a contract for which the secretary is fiscally responsible. Beyond this fiscal control, however, the contract should follow the specifications set out by EAC. This is clearly the intent of the Alexander-James Study Group.

We do not think the EAC should be launched without resolution of this issue. Assertions of goodwill, while perhaps adequate in the short term, only defer the problem for a time. Past history provides compelling lessons in this regard.

ASSEMBLING THE BEST AND BRIGHTEST

Every effort should be made to engage the best of our nation's talent in the membership of the council, in its executive director, its staff, and its various panels.

In addition to the question of the EAC's independence, attention must also be paid to the mechanisms set in place for the selection of staff and the recruitment of panel members. Because NAEP policy will be developed by these individuals, every effort should be made to ensure that a breadth of perspectives is represented among the staff and its panels. Efforts should also be made to attract senior fellows from the

field to form a significant part of the EAC staff. Some positions at least should be designated as visiting appointments to be held for a limited term, such as two years. It would be seriously detrimental to the EAC if it were without regular access to the best research in our universities and the best educational practice in our schools. We recommend that formal arrangements for the appointment and support of senior fellows be incorporated into the staffing plan for the EAC.

NAEP AND THE EDUCATIONAL RESEARCH COMMUNITY

The national assessment is a large applied research and development effort. Our ability to conduct such an assessment draws on a broad base of developments in educational research and related disciplines. If NAEP is to play a larger role in school improvement efforts, then much more attention must be given to the intellectual connections that support, nurture, and sustain it.

As we begin to understand better the prerequisites for an effective NAEP, there is a natural tendency to embed a variety of research and development activities within the NAEP organization. The background papers are replete with recommendations that NAEP be expanded to assume responsibility for a broad array of such activities. A large movement in this direction should not be encouraged. To be sure, both the panel and the background papers are correct in arguing that major research and development activities are needed to advance techniques of measurement and to develop fuller knowledge of the instructional causes of variation in student performance. But the allocation of responsibility for pursuing these various problem areas requires more consideration.

The new NAEP, with its very full agenda, is in danger of spreading its efforts too widely and thereby sacrificing quality in its principal activities. From this point of view, it is simply not efficient to centralize all related research activities under NAEP. Nor is it scientifically sensible to centralize all research activities within a government agency. Without the interactions of a vital external research community, agency efforts can take on an insular self-fulfilling quality that does not serve the public well. A healthy scientific enterprise requires a commitment to public activities that foster a critical exchange of ideas and evidence among opposing views. Through such engagements knowledge advances, and the objectivity of the overall enterprise is safeguarded. These concerns are particularly salient in applied contexts, such as education, where strongly held interests and beliefs can come into play.

DESCRIPTIVE REPORTING SCHEMES

We recommend that, to the maximal extent technically feasible, NAEP use descriptive classifications as its principal reporting scheme in future assessments. For each content area NAEP should articulate clear descriptions of performance levels, descriptions that might be analogous to such craft rankings as novice, journeyman, highly competent, and expert. Descriptions of this kind would be extremely useful to educators, parents, legislators, and an informed public.

As NAEP continues to embody new technical advances in measurement theory, there is a real danger of getting lost in the numbers. For example, the major headings employed in the literacy report are scale score categories ranging from 150 to 400 in increments of 50 or 75. These numbers are arbitrary from both a substantive and technical point of view. Any range of values could have been employed. There is a danger of misuse of numbers like these by well-meaning policymakers who have little or no sense of their limitations.

A great deal of test data is difficult to interpret. What does a level 400 on a reading test mean? Such scores can be used for comparison across time and localities, but the nation's report card would be more broadly informative if it provided clear descriptions of the levels of competence demonstrated by our children. Much more important than scale scores is the reporting of the proportions of individuals in various categories of mastery at specific ages. In several fields, particularly reading and mathematics, we are in a position to describe beginning, average, and advanced competence at various ages. In other areas, such as writing, science, and computer literacy, research remains to be done. NAEP efforts in this area can profit both from the current endeavors of subject-matter specialists and from scientific advances in understanding student learning and cognitive skills. NAEP has already made progress in this direction, and we encourage further effort.

MONITORING THE CONSEQUENCES OF A NEW NAEP

As the mission of NAEP expands, the possibility of problems arising from misinterpretation of data and overzealous use of test results may grow. We recommend that NAEP include provision for an independent monitoring of its impact on schools to assure that it is doing the best job possible of informing meaningful school improvement efforts.

Assuming that the new NAEP is structured along the lines recommended by the Study Group, we feel it is essential to build into the plan

provisions for a group, outside of NAEP itself, to monitor NAEP's impact. Although school practice has been largely impervious to NAEP activity in the past, this may change dramatically; and as we have already suggested, not all these changes may be for the best. In a large complex undertaking such as the new NAEP, there are ample opportunities for misunderstanding, and the list of troublesome problems that could arise is long. Teachers may begin to feel pressure to reduce emphasis on topics that NAEP does not cover so as to improve their showing on the subset of school goals that NAEP does assess. NAEP testing may act indirectly to restrict the course offerings available both to weaker and to gifted students, and over time textbook design may increasingly resemble NAEP test-item design. We are not predicting that any of these problems will necessarily emerge from a new NAEP, but other testing experience shows that such events are possible. It would not be prudent to assume that the new NAEP will be free of unintended consequences.

CONCERN ABOUT STATE-BY-STATE COMPARISONS

We are concerned about the emphasis in the Alexander-James report on state-by-state comparisons of average test scores. Many factors influence the relative rankings of states, districts, and schools. Simple comparisons are ripe for abuse and are unlikely to inform meaningful school improvement efforts.

State average scores on tests like the SAT have been much misused. Although the sampling technique proposed for NAEP will obviate many of these abuses, the ability of a state or locality to examine its progress over time is much more informative than the comparison with other states or localities at any one point in time. Because of the many variables contributing to the diversity of our educational institutions, among states and among localities, the simple ranking of geographic units by achievement level is rarely informative. Not surprisingly, schools with greater resources and fewer problem students routinely fill the upper ranks. So what have we learned?

To say anything about the efficiency of schools would require adjustments for the substantial variations in the "raw materials" with which a school system begins its work and the factors in the home and community that support or impede teachers' efforts. A variety of statistical adjustments can be attempted to compensate for these basic differences among localities, but even when carefully carried out, these analyses provide very little information for school improvement. Similar concerns apply to international comparisons. Cultural differences across

nations make reasonable interpretations of these data very difficult. In-depth studies of considerable extent are required for adequate analysis and interpretation.

A NEED FOR MORE REFLECTION

The issue of setting priorities in the face of competing opportunities for expenditures and constrained federal budgets is not considered in the full report of the Alexander-James panel. This issue can be addressed constructively only when alternative ways of spending federal funds are evaluated along with the likelihood of their improving education.

In particular we wonder about the relative importance of the recommendation that the federal government assume the full cost of collecting data and reporting results separately for each state. This is the single most expensive recommendation, constituting almost 35 percent of the proposed new NAEP annual budget of $26 million.

The most obvious reason for changing NAEP design to permit the reporting of individual state results is that federal statistics are generally published in this way. Indeed, if it were not for political concerns prominent in the 1960s about federal intrusion into local educational decision making, NAEP would almost surely have been designed that way in the beginning. Since the majority of chief state school officers now support a separate reporting by state, there is no reason not to proceed.

We would be concerned, however, if a precipitous decision were made to extend the data–collection plan to permit state-by-state comparisons, only then to find that a lack of funds precluded NAEP from pursuing anything else. In general, there are many ways resources could be deployed to help inform school improvement efforts. In an environment forced to trade off among competing goods, these options merit careful consideration.

For example, as we noted earlier, the basic NAEP design is not very useful for examining the effects of instructional variables on school outcomes. One option would be to support smaller, more intensive assessments specifically designed to address these instructional questions. This would permit a greater understanding of the variation among districts within states and among schools within districts. NAEP might select for further study specific schools that represent important pedagogical concerns—for example, institutions that are particularly strong in science and mathematics for girls and minorities, or schools that accelerate learning for gifted children. In general, there are many opportunities

where in-depth study of carefully selected sites would enable us to learn much about possible school improvements. Taking this approach, NAEP would provide not only an annual reporting on "how we are doing" but also a constant flow of new knowledge on how various aspects of the schooling process and organization contribute to learning.

The Alexander-James report also makes important recommendations concerning expansion of the subject-matter domains covered by the national assessment and using advances in the art of assessment. The importance of educational assessment as an accountability device seems likely to grow in the years ahead. We have already witnessed many school problems precipitated by testing programs that were poorly conceived and implemented. Helping states and individual districts avoid these pitfalls could well be one of NAEP's most significant contributions.

The Alexander-James panel has identified many possible extensions of NAEP. Should the federal government be unable, however, to support the full range of recommended activities, then one option meriting consideration would be to establish a state/federal matching funds arrangement. This arrangement would encourage individual states to take part, yet create sufficient resources to support both the full assessment and related research and development activities.

In concluding, the Academy Committee commends the Alexander-James Study Group for considering an enormous range of issues in a very short period of time. In a necessarily hurried report, they were unable to subject many of the issues raised in the commissioned papers to a full cross-current of informed criticism. Although this intense examination still needs to be done, the Study Group has performed a remarkable job of identifying possible contributions to an evolving and augmented NAEP effort. We hope that our commentary, also hurried and far from exhaustive, adds to their good work and that the two taken together may provide the secretary and the public with useful advice in setting the future course for the nation's report card.

Appendixes

Appendix A

THE SUBGROUPS

POLICY, GOVERNANCE, AND IMPLEMENTATION SECTION 1

Convener: Francis Keppel*
Harvard University

Gordon M. Ambach
Commissioner of Education, New York

Richard A. Boyd
State Superintendent of Education, Mississippi

Joseph M. Cronin
Massachusetts Higher Education Assistance Corporation

Steven F. Ferrara*
Maryland State Department of Education

Jerome T. Murphy
Harvard University

Ramsay Selden
Council of Chief State School Officers, State Education Assessment Center

Richard C. Wallace
Superintendent of Schools, Pittsburgh, Pennsylvania

Frank Walter
Ohio Department of Education

*Author or coauthor of a commissioned paper. See Appendix B.

POLICY, GOVERNANCE, AND IMPLEMENTATION SECTION 2

Convener: Michael W. Kirst*
Stanford University

Dale Carlson
California State Department of Education

Emerson Elliott
OERI, United States Department of Education

Pascal D. Forgione, Jr.
Office of Research and Evaluation, Connecticut State Department of Education

Edward H. Haertel*
Stanford University

Bill Honig
California State Superintendent of Public Instruction

Alexander I. Law
California State Department of Education

Gil Im Obersteg
Federal Liaison Office, California State Department of Education

Paul Sandofar
Office of Research, South Carolina Department of Education

David E. Wiley*
Northwestern University

INTERNATIONAL EDUCATION TESTING

Convener: Michael Timpane
Teachers College, Columbia University

Willard J. Jacobson
Teachers College, Columbia University

Richard M. Wolf*
Teachers College, Columbia University

COGNITIVE SKILLS ASSESSMENT

Convener: George A. Miller*
Princeton University

Susan F. Chipman*
Office of Naval Research

Sandra P. Marshall*
San Diego State University

Samuel Messick
Educational Testing Service

Raymond S. Nickerson*
Bolt, Beranek, and Newman

Lauren B. Resnick*
University of Pittsburgh

READING ASSESSMENT

Convener: Jeanne S. Chall*
Harvard University

John B. Carroll*
University of North Carolina at Chapel Hill

Mary E. Curtis*
Harvard University

Wood Smethurst*
Emory University

Richard L. Venezky*
University of Delaware

Joanna P. Williams*
Columbia University

WRITING ASSESSMENT

Convener: Eva L. Baker
University of California, Los Angeles

Joan Boykoff Baron
Connecticut Assessment of Educational Progress

William E. Coffman*
University of Iowa

Sarah Freedman
University of California, Los Angeles

Edys Quellmalz*
University of California, Los Angeles

Paul L. Williams*
CTB/McGraw-Hill

SOCIAL STUDIES ASSESSMENT

Convener: Jack R. Fraenkel*
San Francisco State University

Ken Carlson*
Rutgers University

Catherine Cornbleth*
University of Pittsburgh

Alan L. Lockwood
University of Wisconsin

Fred M. Newmann*
University of Wisconsin—Madison

Stuart B. Palonsky
University of Missouri

James P. Shaver*
Utah State University

DESIGN AND STRUCTURE

Convener: David E. Wiley*
Northwestern University

R. Darrell Bock*
University of Chicago

George E. Hall
Baseline Data Corporation

Annegret Harnischfeger
Northwestern University

Larry V. Hedges
University of Chicago

Calvin C. Jones*
NORC

Lyle V. Jones*
University of North Carolina at Chapel Hill

C. Philip Kearney*
University of Michigan

Mark D. Reckase*
American College Testing Service

Bruce D. Spencer*
Northwestern University

COSTS

Convener: Joseph M. Cronin*
Massachusetts Higher Education Assistance Corporation

Emerson Elliott
OERI, United States Department of Education

Calvin C. Jones*
NORC

Douglas Rhodes
Budget Office, NAEP

Ramsay Selden
Council of Chief State School Officers, State Education Assessment Center

Rene Sloboski
Westat

Appendix B

THE COMMISSIONED PAPERS

Bloom, Benjamin S. The Home Environment and School Learning.

Bock, R. Darrell. Designing the National Assessment of Educational Progress to Serve a Wider Community of Users.

Carlson, Ken. The National Assessment of Educational Progress in Social Studies.

Carroll, John B. Scales and Other Problems in the NAEP Reading Assessment: Critical Comments.

Chall, Jeanne S. School and Teacher Factors and the NAEP Reading Assessments.

Chipman, Susan F. What Is Meant by "Higher-Order Cognitive Skills"?

Coffman, William E. Recommendations on Writing Assessments for Future NAEPs.

Cohen, Michael. Designing State Assessment Systems.

Cornbleth, Catherine. Assessing Skills and Thinking in Social Studies.

Cronin, Joseph M. The Cost of National and State Educational Assessments.

Curtis, Mary E. The National Assessment of Reading: Past and Future Directions.

Davis, Josephine D. Procedures for Improving the NAEP Mathematics Assessment of Black Youth.

Ferrara, Steven F., and Thornton, Stephen J. Using NAEP for State-by-State Comparisons: The Beginnings of a "National Achievement Test" and "National Curriculum."

Fraenkel, Jack R. The Assessment of Social Studies Knowledge.

Gardner, Howard, and Grunbaum, Judith. The Assessment of Artistic Thinking: Comments on the National Assessment of Educational Progress and the Arts.

Guthrie, John T. Roles of the National Assessment of Educational Progress in International Studies.

Note: Since these papers are an extremely valuable by-product of our study, we have placed them in the ERIC system, to be preserved and made available to anyone interested. Requests for commissioned papers should be addressed to ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, Princeton, New Jersey 08541.

Haertel, Edward H. Domain Definition and Exercise Generation as Functions of the National Assessment of Educational Progress.

Haney, Walt, and Madaus, George. Effects of Standardized Testing and the Future of the National Assessment of Educational Progress.

Hathaway, Walter E. Toward an Ideal System of National Assessment of Educational Progress: A Local Perspective.

Herman, Joan L. What Do the Test Scores Really Mean? Critical Issues in Test Design.

Jones, Calvin C. Relationships between the National Assessment of Educational Progress and the National Educational Longitudinal Studies Program.

Jones, Lyle V. The Future Assessment of Educational Progress: Specifying Background Variables and Subpopulations.

Kearney, C. Philip. NAEP: A National Data System for the Twenty-first Century.

Kirst, Michael W. Roles, Governance, and Multiple Uses for a New NAEP.

Lewis, Anne E. The Politics of Testing/Assessment; or, A Chameleon in the Classroom.

Marshall, Sandra P. Understanding Arithmetic Story Problems.

Miller, George A. Explanatory Skills.

Newmann, Fred M. The Assessment of Discourse in Social Studies.

Nickerson, Raymond S. Reasoning and Argument Evaluation.

Peterson, Paul E. Purposes of the National Assessment for Educational Progress.

Quellmalz, Edys. Recommendations for the Design of the NAEP Writing Tasks.

Reckase, Mark D. Position Paper on the Potential Use of Computerized Testing Procedures for the National Assessment of Educational Progress.

Resnick, Daniel P., and Resnick, Lauren B. Understanding Achievement and Acting to Produce It: Some Recommendations for NAEP.

Rivera, Charlene. The National Assessment of Educational Progress: Issues and Concerns for the Assessment of Hispanic Students.

Romberg, Thomas A. National Assessment of Mathematical Performance.

Shaver, James P. National Assessment of Values and Attitudes for Social Studies.

Smethurst, Wood. Early Beginnings, Success and Failure in Teaching Young Children to Read: Some Abiding Questions and Intriguing Possibilities.

Spencer, Bruce D. Efficient Methods for Sampling Out-of-School Seventeen-Year-Olds in the National Assessment of Educational Progress.

Sticht, Thomas G. Issues in Indexing Functional Literacy.

Tyler, Ralph W. The Governance of the National Assessment of Educational Progress.

Venezky, Richard L. Literacy and the NAEP Reading Assessments.

Welch, Wayne W. Some Thoughts on the National Assessment of Educational Progress in Science.

Wiley, David E. Use and Usability of National Assessment Information.

Williams, Joanna P. Assessment at Age Seven.

Williams, Paul L. NAEP Writing Assessment: A Committee Member's Perspective.

Wolf, Richard M. NAEP and International Comparisons.

Appendix C

PUBLIC LAW 98-511, SECTION 405(e)

(1) In addition to the other responsibilities of the Office under this section, the Office shall carry out, by grant or cooperative agreement with a nonprofit educational organization, a National Assessment of Educational Progress which shall have as a primary purpose the assessment of the performance of children and young adults in the basic skills of reading, mathematics, communication, and other subjects and skills. Such a National Assessment shall—

(A) collect and report at least once every five years data assessing the performance of students at various age or grade levels in each of the areas of reading, writing, and mathematics;

(B) report periodically data on changes in knowledge and skills of such students over a period of time;

(C) conduct special assessments of other educational areas, as the need for additional national information arises;

(D) include in assessment activities information on special groups of individuals;

(E) provide technical assistance to State educational agencies and to local educational agencies on the use of National Assessment objectives, primarily pertaining to—

(i) the basic skills of reading, mathematics, and communication, and

(ii) on making comparisons of such assessments with the national profile (including special population profiles) and change data developed by the National Assessment; and

(F) with respect to each State which voluntarily participates in accordance with paragraph (5), provide a statement of information collected by the National Assessment for each such State.

(2)(A) The organization through which the Office carries out the National Assessment shall be responsible for overall management of the National Assessment. Such organization shall delegate authority to design and supervise the conduct of the National Assessment to an Assessment Policy Committee, established by such organization. The Assessment Policy Committee shall be composed of—

(i) five members appointed by the organization of whom two members shall be representatives of business and industry and three members shall be representatives of the general public; and

(ii) fourteen members appointed by the organization from the categories of membership specified in subparagraph (B).

(B) Members of the Assessment Policy Committee appointed in accordance with subparagraph (A)(ii) shall be—

"(i) one chief State school officer;

"(ii) two State legislators;

"(iii) two school district superintendents;

(iv) one member of a State board of education;

(v) one member of a local school board;

(vi) one Governor of a State;

(vii) four classroom teachers;

(viii) one elementary school principal; and

(ix) one secondary school principal.

(C) The Assistant Secretary shall serve as an ex officio member of the Assessment Policy Committee. The Assistant Secretary shall also appoint a member of the Council to serve as nonvoting member of the Assessment Policy Committee.

(D) Members appointed in accordance with subparagraph (A) (i) and (ii) shall be appointed for terms for three years on a staggered basis.

(3) The Assessment Policy Committee established by paragraph (2) shall be responsible for the design of the National Assessment, including the selection of the learning areas to be assessed, the development and selection of goal statements and assessment items, the assessment methodology, the form and content of the reporting and dissemination of assessment results, and studies to evaluate and improve the form and utilization of the National Assessment. The appropriateness of all cognitive, background, and attitude items developed as part of the National Assessment shall be the responsibility of the Assessment Policy Committee. Such items shall be subject to review by the Department of Education and the Office of Management and Budget for a single period of not more than sixty days.

(4) Each learning area assessment shall have goal statements devised through a national consensus approach, providing for active participation of teachers, curriculum specialists, subject matter specialists, local school administrators, parents, and members of the general public. All items selected for use in the assessment shall be reviewed to exclude items which might reflect racial, sex, cultural, or regional bias.

(5) Participation in the National Assessment by State and local education agencies selected as part of a sample of such agencies shall be voluntary.

(6) The Secretary shall provide for a periodic review of the National Assessment. This review shall provide an opportunity for public comment on the conduct and usefulness of the National Assessment and shall result in a report to the Congress, the President, and the Nation on the findings and recommendations, if any, of the review. The Secretary shall consider the findings and recommendations in designing the competition to select the organization through which the Office carries out the National Assessment.

Appendix D

A CHRONOLOGY OF THE NATIONAL ASSESSMENT

1963	U.S. Commissioner of Education Francis Keppel requests Ralph Tyler to develop an assessment program
1964-68	Assessment is designed and pretested
1969	First assessment conducted: writing, science, and citizenship assessed
1971	Reading and literature assessed
1972	Social studies and music assessed
1973	Mathematics and science assessed
1974	Writing and career and occupational development assessed
1975	Reading and art assessed
1976	Citizenship and social studies assessed
1977	Science assessed
1979	Writing, art, and music assessed
1980	Reading and literature assessed
1982	Mathematics, citizenship, and social studies assessed
1984	Reading and writing assessed
1986	Reading, mathematics, science, and computer competence assessed
1988	Proposed that reading, writing, citizenship, history, and geography be assessed

For all of the content areas given above, information on nine-, thirteen-, and seventeen-year-olds across the nation has been collected. It should be noted that some assessments were conducted during a school year and thus spanned two calendar years. For convenience, the above listing shows only the second year of the two.

Appendix E

THE ELEMENTARY/SECONDARY INFORMATION DATA SYSTEM

The federal Center for Education Statistics is charged with the statutory responsibility of collecting data on the state of education in the United States. The center meets that responsibility for elementary and secondary education by its management of the Elementary/Secondary Information Data System (ESIDS), a data bank with national access. The center has recently redesigned the system and is currently enlarging it with new survey data. When completed, ESIDS will form one integrated system consisting of two types of components.

UNIVERSE DATA

Public School Districts: A school district census (identification and type)
Public and Private Schools: A census of all public and private schools (identification, enrollment, staffing, and type)
State Aggregate Fiscal Data: Revenues, expenditures, and average daily attendance (ADA)
State Aggregate Nonfiscal Data: High school graduates, enrollment by grade, instructional and noninstructional staff
Early Estimates: Universe component (new to the system)

SAMPLE DATA

Sample data illustrate characteristics of each of the following.

Public School Districts
Public and Private Schools
Public and Private School Teachers
Public and Private School Libraries
Public and Private School Administrators (new to system)
Parents of NAEP Students (new to system)
Student Performance (NAEP)
Student Progress over Time (Longitudinal Studies)
Public School Finance (under development)

Teacher Leavers (new to system)
Policies and Practices (new to system)
Early Estimates: Sample component (new to system)

Data will be collected both through state administrative records and through linked surveys. The periodicity will range from annual for universe data, biennial for school and staff characteristics data, to every three or four years for data on teacher leavers and detailed school finance. Data on teacher demand and shortages (based on LEA surveys, school surveys, administrator surveys, and teacher surveys) will be available every two years and will be state-representative for public schools. School and staffing surveys will be implemented fully in 1988; the teacher leaver survey in 1989; and the linkage of process and outcome data in 1990.