

# EVALUATION STUDIES

## THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

**Frank B. Womer**

*University of Michigan*

and

**Wayne H. Martin**

*National Assessment of Educational Progress,  
Denver, Colorado*

The purpose of this paper is to provide the reader with a knowledge of the National Assessment of Educational Progress (NAEP) project. The first section of the paper provides an overview and general discussion of National Assessment; the second section presents some of the operating issues related to the project and its purposes; and the third section provides a short history of the project and its governance.

### AN OVERVIEW OF NATIONAL ASSESSMENT

The National Assessment of Educational Progress was established to monitor the progress of educational attainments in the United States. More specifically, the program was designed to measure the knowledge, skills, and attitudes possessed by young Americans aged 9, 13, 17, and 26 to 35 and to measure changes (growth or decline) in their educational attainments over time. To accomplish this, assessments in one or more learning areas are conducted annually at the three in-school ages and periodically for the young adults (26-35). Table 1 presents the assessment timetable for the first ten years of data collection. The project has conducted the initial or baseline assessment in each of the ten learning areas (art, career and occupational development, citizenship, literature, mathematics, music, reading, science, social studies, and writing); by early 1978, it had completed the second assessments of science, writing, citizenship, reading, social studies, and mathematics and the third assessment of science.

National Assessment also recognizes that special needs for data not included in the ongoing assessment frequently arise. To address such needs, special purpose "probes" are used. A probe is a small scale assessment of a specific topic and is usually administered to only one age group. Intended to provide a "snapshot" of the topic, these probes are not necessarily designed to measure change.

Probes have been administered in the areas of basic life skills and consumerism for 17-year-olds and in the areas of health and energy for adults. In addition, NAEP administered, analyzed, and reported the assessment of functional literacy of 17-year-old students under contract to the Right-to-Read program. Probes provide National Assessment with the flexibility to address timely educational topics without disrupting its program of monitoring the nation's educational progress.

TABLE 1: *Assessment Timetable (Expressed in school years, July through June)*

Learning Area	Baseline Assessment	First Measurement of Change	Second Measurement of Change
Science	1969-70	1972-73	1976-77
Writing	1969-70	1973-74	1978-79
Citizenship	1969-70	1975-76	-
Reading	1970-71	1974-75	-
Literature	1970-71	-	-
Social studies	1971-72	1975-76	-
Music	1971-72	1978-79	-
Mathematics	1972-73	1977-78	-
Career & occupational development	1973-74	-	-
Art	1974-75	1978-79	-

### Objectives Development

One of the distinguishing features of National Assessment is that it is objective-referenced. It was a pioneer in the use of this approach in America. First, objectives are developed, reviewed, and redeveloped employing a consensus approach. Large numbers of national educational leaders are brought together in panels to critique the objectives. Each particular objective retained for an assessment must be judged as important for students to know or be able to do by scholars in the field of that learning area, educators responsible for transmitting that learning area to students, and concerned lay people. Use of this approach is designed to eliminate fads from the assessment of a learning area. The review and redevelopment of objectives takes place several years prior to assessment of a learning area. This ensures that new approaches and emphases are considered for each assessment while also maintaining considerable continuity of objectives over time. The National Assessment objectives in each learning area are printed and are made available nationally though they were not developed for that purpose. They may well represent the closest thing that the U.S. has to a statement of national goals and objectives in education. Since the objectives must be written to satisfy groups of individuals with somewhat diverse points of view, they do not include much that is controversial.

### Item Development

The items or exercises to be used for National Assessment are developed specifically as measures of the assessment objectives in each learning area. They are written by a variety of item writers who, for the most part, are educators or other learning-area specialists. They are reviewed by individuals representative of the same major groups used in the development of the objectives, i.e., scholars, educators, and lay people. As well as judging the content validity of each item, reviewers are asked to evaluate items on the basis of stereotyping of or offensiveness to

racial, ethnic, religious, political, social, and regional groups and organizations.

Items are tried out on small groups of students in order to eliminate specific problems of wording. Item analyses are used in an effort to locate problems, but items are *not* subject to the usual item analytic procedures designed to maximize item discrimination and to secure item difficulties as close to 50% as possible. Instead, sets of items are developed with a wide range of item difficulties. The reason for this is based on a desire to be able to report the results of individual items in a scheme that provides information about what "most" students know or can do, what "typical" students know or can do, and what the "ablest" students know or can do.

### Data Collection

A national probability sample of students is assessed each year for the in-school age groups; similarly a national probability adult sample is used for each periodic adult assessment. Students are reached through a sample of school buildings while a sample of households is used for the adult assessments.

To minimize its impact upon individual school districts, NAEP draws a four-year master sample of schools. This allows the project to guarantee a school district that if they are selected for participation in the sample, it will only be once during a four-year period for the vast majority of school districts. Although it is necessary to include a portion of the school districts within the nation's 12 largest standard metropolitan sampling areas in each year's assessment to guarantee a valid national sample, the sample is designed so that a particular school within these districts will only be selected once in a four-year period. Even though participation is voluntary, since NAEP is not a direct arm of the government and has no legal grounds for requiring participation, the status of the project is such that over 90% of schools selected do participate.

About 10-12 packages are administered in each assessment. A "package" is a set of objective-referenced questions. Each student takes only one package, requiring a maximum of 50 minutes. A student may opt not to take a test if he/she wishes. Most packages are administered to groups of 16-20 students using a paced-tape technique. The administrator plays a recorded set of directions, with each question (including distractors of multiple-choice questions) read aloud on the tape. The tape provides silence for a period of time sufficient to answer each question. This technique is not used for the reading assessments nor for individually administered items (a few packages in each assessment include items which, because of their nature, require individual administration rather than the usual paced tape). These procedures result in minimum disruption to local school staffs and officials.

In order to further guarantee uniformity of data collection, each assessment is administered by trained personnel. Permanent district supervisors stationed all over the country have the responsibility for handling local arrangements for NAEP. They do some of the package administration themselves and hire and train local persons for the remainder. Thus, a school district is asked only to provide the students selected in the sample and space for test administration.

Respondent names are not recorded on the packages. They are used only for selection of the local sample. Thus, students and adults are guaranteed anonymity.

### Reporting

Results are reported for the nation as a whole and for the four regions of the country, as well as breakdowns according to sex, race, size-and-type of community, and level of parental education for each age group. No comparisons are made among individual students, schools, school districts, or states; the sample design does not permit making inferences about any geographic unit smaller than a region. The project reports results on all items and, in addition, publishes the entire text of approximately half of the items from each assessment. This allows for a detailed understanding of the assessment's content and purposes and for great flexibility in interpretation or further analyses. The unpublished items are kept secure and used in subsequent assessments of the learning area to measure change over time.

National Assessment does *not* itself interpret the assessment data. Early in the project there were many fears expressed that NAEP was a move in the direction of establishing a national curriculum. In the U.S., local autonomy in curriculum development is a cherished concept. However, experience with the first few reports revealed that numbers and statistics alone do not constitute a report that will be widely used by the public, policy makers, or even practitioners in the field. To attempt to meet the needs of its various audiences, a variety of reports for each assessment are developed now. Selective reports (dealing with a subset of the items) and an overview of each assessment are prepared for the public, policy makers, and practitioners. Technical reports are prepared for the research and evaluation units of state and local education agencies and the educational research community. A brochure which provides a nontechnical summary of each assessment is also prepared for general distribution.

National Assessment also takes the initiative to involve representatives from various professional organizations in the development and review of objectives and items, the review of analysis and reporting plans, and the actual review and evaluation of assessment results. Their comments and reactions are often included in special interpretation chapters of project reports. The early concerns about "influences" from a national project seem to have changed enough so that the interpretations developed by the various professional groups independent of NAEP are welcomed by the educational community.

### Utilization of Results

Documentation of a specific national educational policy determination based upon direct evidence from National Assessment results cannot be made. In the U.S., educational decision-making is so diffuse and the amount of information available is so great that it is difficult to relate specific decisions to specific inputs. Probably many American educators would claim that the well-known Coleman report has influenced decisions in the area of attempting to equalize educational opportunities across various minority groups in this society. But, direct cause and effect would be difficult to establish if one keeps in mind the significant societal

changes taking place at the time that study was done.

In like fashion, National Assessment results are being quoted by an increasing number of educators who testify before congressional committees, who make speeches at professional meetings, who write textbooks and journal articles, and so on. National Assessment is providing, for the first time, a regular ongoing data base about the specific accomplishments of American youth. The first change information from NAEP has generally shown declines in the levels of attainment in science knowledge, in writing skills, and in political knowledge at all ages. The picture in reading is not as clear; lower-achieving students and younger students have shown gain, while, for older students, a decline was observed. These data are being widely quoted in the public press as well as in educational journals and may very well be contributing to a general public concern about declining academic achievements. However, causal relationships can only be speculative at this time.

Within the educational community, NAEP's acceptance is so widespread that its results now are quoted with a mere reference to their source rather than with the reference *plus* a short explanation of the project. National Assessment is slowly but surely becoming the "census" of educational attainments nationally, the first source for information about what young Americans know and can do.

### Services

The objectives and items released in National Assessment reports are made available to state and local education agencies and the research community for use in other assessment or evaluation projects.<sup>1</sup> To facilitate the use of NAEP materials and methods, National Assessment has staff who provide service to the project's constituents, which include: state and local education agencies, federal agencies and panels, professional organizations, the education research community, and the general research community. NAEP sponsors an annual large-scale assessment conference which provides a forum for the discussion of policy and technical issues facing assessment/evaluation projects and for the sharing of ideas and technology.

National Assessment conducts regional meetings with a variety of representatives from the state, local, federal, and national levels to review future plans and possible new directions for the project, to consider additional learning areas for assessment, to identify special probes, and to critique current assessment procedures and policies. Mail surveys are also conducted to gather input from constituents. This input is shared with NAEP's two advisory committees and the National Assessment Policy Committee, the governing body responsible for the establishment of policies and guidelines for the project.

For measurement and evaluation specialists, it is this service outreach of NAEP that is its most visible contribution. The evaluator in a state department of education or in a local school district looks to Na-

---

<sup>1</sup>National Assessment prints and distributes a great variety of reports and other documents; a complete listing of all publications can be obtained from: National Assessment of Educational Progress, Education Commission of the States, Suite 700, 1860 Lincoln St., Denver, Colorado 80295, U.S.A.

tional Assessment as one of the prime sources of materials and services. In addition, methodological contributions of NAEP are being replicated in a great many other assessment projects.

### SOME SPECIAL ASPECTS OF THE PROJECT

#### Internal Operation of the Project

As the Exploratory Committee on Assessing the Progress of Education (ECAPE) and its successor, the Committee on Assessing the Progress of Education (CAPE), addressed the task of designing, developing, and initiating a periodic national assessment, it became obvious that the technology required for a large-scale, long-range, objective-referenced undertaking, such as NAEP, did not totally exist. Thus, it was necessary to adapt and develop the technology as the project's design became more concrete and definite. During the development period and the initial years of National Assessment data collection, many activities were being accomplished concurrently and almost independently of each other. This was necessitated by time limitations and the limited staffing of the project. This situation was made more complex by the general inexperience of educational researchers in implementing a project of the magnitude of National Assessment.

In the development years, it appeared very sensible that the logical phases or activities of the project could be handled as separate entities and dealt with when the need arose. It became evident as the project grew and more data were collected that this model was neither efficient nor ideal. While the assessment of a learning area can be divided into logical phases (objectives-development and review, item-development and review, data collection, scoring, analysis, reporting, and service to users), the activities are so interrelated that it is not practical to consider doing each independent of the others. Objectives-development and review and item-development and review are obviously related for an objective-referenced assessment; these two phases produce the materials to be assessed. Yet these activities also have major implications for data collection, scoring, analysis, reporting, and service to users. The assessment activities are basically a chain, with each of the major phases linked to the other phases. Thus, to achieve the integrated assessment of a learning area, National Assessment employs a team concept. Each learning area team includes staff members with major responsibility for different phases of the assessment, but all team members are involved in all phases of the assessment for a particular assessment year. This is done to ensure that the project will develop compatible with data collection, scoring, analysis, and reporting procedures, and that the reports from an assessment will be related to the objectives established for the assessment of the learning area. Because of the long-range nature of National Assessment, it is necessary to begin work on each assessment approximately two years before the beginning of data collection.

Data collection takes place during the school year. Thirteen-year-olds are assessed during October through mid-December; 9-year-olds during January and February; and 17-year-olds are assessed from March through mid-May. Each separate package is taken by about 2,000 students. Thus, a total of about 100,000 students are assessed each year in the NAEP

sample. The scoring of assessment items begins shortly after the starting of data collection for an age group. For most learning areas, analysis and reporting are completed within one year from the end of the data collection for the assessment.

To preserve the continuity of National Assessment, it is necessary for the staff to be working on several different assessment years during any given fiscal year. Table 2 illustrates this by showing the major activities being conducted for various assessment years during fiscal years (FY) 1976, 1977, and 1978. Please note that assessment years are expressed in terms of a school year (July through June), while NAEP's fiscal year is January through December as determined by its funding source.

The operation of the project is handled in part by a permanent in-house staff and in part under contracts to several other organizations. The development of the first specific statements of learning objectives and of the associated items originally was done by contracting with four different organizations. NAEP staff served as monitors of the reviewing process and, with consultant help, made all final selections of objectives and items. In recent years, the developmental activities have been handled in-house. The former point of view was that it was better for staff to remain divorced from the development of the objectives and items in order to maximize objectivity in the review process. The present point of view is that staff can manage both activities without personal involvement or loss of objectivity.

The development of the sample design and the actual data collection are contracted out to a survey organization that specializes in such activities. Test scoring is contracted out to a machine-scoring agency. NAEP maintains a small staff to monitor these operations.

Data analysis and reporting are in-house activities and always have been. A sizeable staff of analysts and report writers has been developed. They are responsible along with an Analysis Advisory Committee for the development of specific analytic techniques appropriate to the National Assessment project and for the preparation of NAEP reports.

### The Measurement of Change

The assessment of change is the prime goal of National Assessment. To determine if the knowledge, skills, and understanding of young Americans have increased, declined, or stayed the same, NAEP readministers, in each assessment, a portion of the items from the previous assessment of that learning area. These items are kept secure and not reproduced in reports of results. Thus, approximately half of the items for each assessment of a learning area are from prior assessments, while the remaining one-half are newly-developed items. This allows the project to measure change over time on an ongoing basis while also allowing it to reflect new direction or emphases in the learning area.

The measurement of change across time places restrictions on a project with regard to flexibility for innovation. At any one point in time, the assessment of a learning area is tied to the data collection procedures from the prior assessment in order to be able to measure change with approximately one-half of the assessment items. While it is possible to implement innovations with the newly-developed items for the assessment,

TABLE 2: Major Project Activities During FY 76, 77, and 78

Learning Areas	Assessment Year	FY 77 Major Activities	FY 77 Major Activities	FY 78 Major Activities
Career & occupational development, Social studies, & Citizenship	1980-81			Objectives review, Item development
Reading & Literature	1979-80		Objectives review, Item development	Item development, Item tryouts
Writing, Music, & Art	1978-79	Objectives review, Item development	Item development, Item tryouts	Item selection, Printing, Sampling <sup>a</sup>
Mathematics	1977-78	Item development, Item tryouts	Item selection, Printing Sampling <sup>a</sup>	Data collection, Scoring
Science	1976-77	Item selection, Printing, Sampling <sup>a</sup>	Data collection, Scoring	Scoring, Analysis
Social studies & Citizenship	1975-76	Data collection, Scoring	Scoring, Analysis	Reporting, Documentation
Reading & Art	1974-75	Scoring, Analysis	Reporting, Documentation	
Writing, Career & occupational development	1973-74	Reporting, Documentation		

<sup>a</sup>The collection of assessment data for 13-year-olds is conducted during October through mid-December of this school year.



those "new" items become the "restrictive" items with which change is measured for the next assessment of the learning area. A similar situation exists for the reporting variables. To report change over time, the background variables must be used in both assessments; new variables may be added in anticipation of their use as reporting variables for future assessments.

For example, assume that parental income rather than level of parental education was now to be used as a reporting variable. It would be necessary for respondents to answer the questions relating to both parental education and parental income during the first assessment of the learning area. By the next assessment of the learning area, it would be possible to delete the questions relating to parental education and only ask the questions pertaining to parental income. While this sounds like a relatively simple change, it must be remembered that students spend less than one hour participating in the assessment. The time that is spent answering additional background questions to allow for changes in reporting variables in effect lessens the amount of time that can be spent responding to the actual assessment items. Thus, National Assessment has chosen to move very cautiously in making changes in its basic data collection plan.

At any time, the option exists to totally modify procedures for any assessment of a learning area. However, the cost of exercising this option is the loss of the ability to measure change over time from previous assessments. This option effectively discards all the previously collected data. The project has never seriously considered this option, as it is in contradiction to the main purpose of the project. National Assessment has, instead, gradually introduced changes and modifications into the assessment.

#### A HISTORY OF THE GOVERNANCE OF NATIONAL ASSESSMENT

By the early 1960s, the annual expenditure of public funds for the formal education of young Americans was approximately \$30-billion.<sup>2</sup> Yet criticism of the educational system abounded. Defenders of the educational establishment found it increasingly difficult to provide evidence that the schools were satisfactorily meeting the educational needs of a modern, technological society. The only readily available measures of educational quality resulting from this public investment of funds were "input" measures such as teacher-student ratios and per-pupil expenditures. The tenuous assumption was made that the quality of educational "outcomes" - what students do or do not know and can or cannot do - was directly related to the quality of the inputs to the educational system. There was no conclusive empirical evidence to support this assumption.

This insufficiency of information became the concern of Francis Keppel, U.S. Commissioner of Education (1962-65). He initiated a series of conferences to explore ways to provide the necessary information. In 1964, as a result of these conferences, John W. Gardner, president of the Carnegie Corporation, asked a distinguished group of educators and other concerned persons to form the Exploratory Committee on Assessing the

---

<sup>2</sup>Based on U.S. Census Bureau data. Actual amounts expended were: 1960, \$24.7-billion; 1962, \$29.4-billion; 1964, \$35.9-billion.

Progress of Education (ECAPE). This committee, chaired by Dr. Ralph W. Tyler, who had been involved since the earliest conferences, was to examine the possibility of conducting an assessment of educational attainments on a national basis.

During this early period in which ECAPE and its successor, the Committee on Assessing the Progress of Education (CAPE), addressed the task of designing, developing, and initiating a national assessment, the Committee was governed by a group of distinguished educators and concerned lay people and financed by foundation funds from the Carnegie Corporation, the Ford Foundation, and the Fund for the Advancement of Education. Although federal officials often participated in meetings and it was assumed that the federal government would eventually fund the project, at no time did the federal government actually assume control of the project.

Shortly after the project began its initial data collection in the spring of 1969 (as CAPE), the funding of the project shifted from foundation funds to federal funds. The federal funding of National Assessment raised many questions and fears about the true intent of the project. Was this an attempt to establish national curricula and/or national performance standards? Was this the first step toward the establishment of eventual federal control over education?

Historically, in the U.S., education has been the constitutional responsibility of state and local governments. A strong tradition of local control of education has developed in this country in contrast to the models espoused by most other countries. At the same time, the involvement of the federal government in education has increased over the years through both legislation and funding. Although people recognized the need for monitoring the nation's educational progress, there was a great deal of concern about the possible dangers inherent in a federally controlled project designed to accomplish this. Legislators, educators, and others were unclear as to whether the value of such a project, directed and controlled by the federal government, outweighed the potential dangers.

In response to these concerns and the voiced opposition to the project because of these concerns, National Assessment was placed under the administration and direction of the Education Commission of the States (ECS), an interstate compact of 46 states, Puerto Rico, and the Virgin Islands. The details of this arrangement were outlined in a Memorandum of Understanding between ECS and the Office of Education, U.S. Department of Health, Education, and Welfare.

The ECS governing board appointed a National Assessment Policy Committee to serve as the governance group for National Assessment and to determine the project's policies and directions. The National Assessment Policy Committee was and continues to be made up of representatives from the political, education, and business communities.

Since 1969, it has been the National Assessment Policy Committee that considers and establishes the policies of the project. The committee, meeting quarterly with the staff, decides upon possible new programs and/or directions and modifications to the existing programs and procedures while at the same time remembering that the main purpose of the project is to measure changes in the nation's educational attainments over time. Perhaps more important, it has maintained the "quasi independence" of the project from the federal government.

All operating funds for the program, about \$5-million a year, come from the U.S. Department of Health, Education, and Welfare. The operation of NAEP is monitored by one of HEW's divisions, the National Center for Educational Statistics. The founders and developers of NAEP knew that continuing financial support for such a project had to come from the federal government. They have made every effort, however, to develop a project that accepted that reality while insisting on the development of a professionally acceptable governance system which up to this point in time has been able to maintain its independence to make program decisions on the basis of educational importance.