DOCUMENT RESUME

ED 480 086 CG 032 659

AUTHOR	Wise, Lauress L.
TITLE	The National Assessment of Educational Progress: What It Tells Educators.
PUB DATE	2003-08-00
NOTE	15p.; In: Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators; see CG 032 608.
PUB TYPE	Information Analyses (070)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Educational Assessment; *Evaluation Methods; Predictive Measurement; *Testing Programs; Trend Analysis
IDENTIFIERS	*National Assessment of Educational Progress

ABSTRACT

Since its inception in 1969, the National Assessment of Educational Progress (NAEP) has been the nation's leading indicator of what American students know and can do. The focus of this chapter is on how NAEP, as it exists today, may be useful to educators, in particular four aspects of NAEP that may be of wide interest and use. This chapter concludes with a discussion of planned or possible enhancements to NAEP that could further increase its usefulness to educators. (Contains 12 references and 2 tables.) (GCP)

The National Assessment of Educational Progress: What It Tells Educators

By
Lauress L. Wise

Chapter 51
The National Assessment of Educational Progress
What It Tells Educators

Lauress L. Wise

On October 4, 1957, the Soviet Union launched the first artificial satellite from the Baikonur cosmodrome in Kazakhstan (see www.batnet.com/mfwright/sputnik.html). America's self-image as the world's technological leader was shattered. In the ensuing years, numerous efforts were launched to improve the education of American youth and thus restore our globaal competitiveness. These efforts ranged from "new math" to Project TALENT, an intensive study of 400,000 students in American high schools in 1960. Amid these efforts, Ralph Tyler pursued the sensible notion that we should regularly assess elementary and secondary student achievement so as to measure the progress of education. Planning conferences were held beginning in 1964, and later in the 1960s the National Assessment of Educational Progress (NAEP) was launched (Jones, 1996). A recent review of NAEP by the National Academy of Education (Glaser, Linn, & Bohrnstedt, 1997) begins with the statement, "Since its inception in 1969, the National Assessment of Educational Progress (NAEP) has been the nation's leading indicator of what American students know and can do" (p. 1)

In its beginning, NAEP reported student performance on specific test questions selected to represent subject areas for students at ages 8, 12, and 17. This reporting process has undergone a number of significant changes over the past 30 years, for example, grade cohorts (i.e., grades 4, 8, and 12) have replaced age cohorts in assessments. In the mid-1980s, item response theory (Lord & Novick, 1968) was introduced to provide an overall score scale as a complement to item-by-item results. In response to a book by Alexander and James (1987), an independent governing board was created to oversee the content and administration of the assessment in partnership with the U.S. Department of Education (Vinovskis, 1998). Beginning in 1990, state results were released along with national trend information. The No Child Left Behind Act, passed

by Congress in 2001, requires state participation in NAEP. NAEP results will likely be used to audit state measures of yearly educational progress.

Currently three relatively distinct components comprise NAEP. National NAEP reports student achievement for the nation as a whole relative to current content frameworks for each subject area. State NAEP reports results for each participating state on a more limited set of subjects and grades. The long-term-trend NAEP reports student results at the national level based on the content and format of assessment that has been common over the last several decades.

A detailed recounting of the history of NAEP is outside the scope of this chapter. The following sources provide much more detailed information on how NAEP has evolved and what changes may lie ahead:

Alexander and James (1987)
Jones (1996)
Glaser et al. (1997)
Pelligrino, Jones, & Mitchell (1999)

The National Center for Education Statistics within the U.S. Department of Education maintains a website that includes a wide range of information on the current NAEP: http://nces.ed.gov/nationsreportcard/.

The focus of this chapter is on how NAEP, as it exists today, may be useful to educators, in particular four aspects of NAEP that may be of wide interest and use. First, NAEP provides content frameworks for particular subjects that reflect a national consensus on what 4th-, 8th-, and 12th-grade students should know and be able to do. Second, the National Assessment Governing Board (NAGB), which Congress created in 1988 to set NAEP policy, has adopted performance standards for each grade and subject indicating Basic, Proficient, and Advanced mastery of the knowledge and skills specified in each of the content frameworks. Third, NAEP has contributed many innovations to the assessment of student achievement, and questions released by NAEP provide concrete examples of these innovations. Finally, NAEP continues to provide national normative data at the test question level as well as for the overall NAEP reporting scales. This chapter concludes with a discussion of planned or possible enhancements to NAEP that could further increase its usefulness to educators.

National Content Frameworks

NAEP has contributed significantly to the dialogue about what we should be teaching students at the elementary, middle, and high school levels. There is of course a rich tradition of state and local control of schools, yet there is also a growing recognition that students will have to compete in a national, if not international, employment market. Thus while emphases may vary, there is surely a core set of skills that students will need in order to succeed in college, the workplace, avocational pursuits, and civic responsibility. Indeed, business and labor have expended extensive effort to define essential workplace skills through the Labor Secretary's Commission on Acquiring Necessary Skills (SCANS) and later the National Skills Standards Boards. The NAEP content frameworks reflect an important effort to identify essential knowledge and skills for students in all states and local districts.

A national consensus process is used. Several features of the content frameworks developed by NAEP make the frameworks noteworthy. First is the careful consensus process used in developing and adopting these frameworks. The NAGB has contracted with the Council of Chief State School Officers and similar broad-based organizations to manage the development of recommended frameworks. Professional organizations that represent content specialists, such as the National Council of Teachers of Mathematics, have played a leadership role in framework development. The NAGB handles the adoption of content frameworks. NAGB is an independent, bipartisan organization chartered by Congress to manage the content and timing of NAEP assessments. By statute, it includes two governors, two state legislators, two chief state school officers, and a mix of district and school personnel, content specialists, measurement experts, and the general public (Vinovskis, 1998). Before approving the frameworks recommended by a development contractor, NAGB holds hearings at locations throughout the nation to obtain public comment on the proposed frameworks. A subcommittee of NAGB members manages these hearings and processes the input, working with the development contractor on potential changes to accommodate suggestions from the hearings. The entire board must approve final frameworks before they are initiated.

The frameworks are inclusive. If significant consequences for students or schools were attached to scores from NAEP, it would be necessary to limit the content of what is tested to material that is taught in all

schools. At the very least, this would mean limiting NAEP content to the intersection of the frameworks adopted by the different states. Because, as of this writing, there are not any direct consequences attached to NAEP scores, this restriction does not apply. In fact, NAEP frameworks tend to be inclusive, encompassing content that is deemed significant by all states and by other sources as well.

The frameworks are forward looking. The NAEP frameworks are not merely a reflection of what is currently taught and are not limited to what is currently included in one or more of the state frameworks. The frameworks attempt to balance what is being taught with expert judgment about what should be taught. In this sense, the frameworks are forward looking and provide a model that many states find useful in updating and revising their own content standards.

What frameworks are available? Table 1 lists the NAEP content frameworks used with recent or pending assessments. In each case, the frameworks specify content for the assessments at the 4th-, 8th-, and 12th-grade levels. A revised framework for mathematics will be used with the 2005 assessment, and a framework for economics is under development. The NAGB website lists updated information: http://nagb.org/. Copies of most of the frameworks can be downloaded from this site. Instructions for ordering printed copies from the NAGB are also available there.

Table 1. NAEP Content Frameworks

Subject	Assessment Years
Mathematics	1996, 2000
Reading	1992–2000
Science	1996, 2000
History	1994, 2001
Geography	1994, 2001
Foreign Language	2003
Writing	1998
Civics	1998
Arts	1997

Student Performance Standards

Since 1990, NAGB has addressed not just what students should know and be able to do as indicated by the content frameworks, but also the level of mastery of each subject that constitutes proficiency. In the beginning, NAEP reported percentages of students answering individual questions correctly. In 1983, the NAEP grant was moved from the Education Commission of the States to the Educational Testing Service (ETS). ETS constructed an overall scale based on item response theory and began reporting yearly means on this scale. The new scale allowed yearly gains to be summarized in terms of a single number rather than reported separately for each test item. Several attempts were made to describe what students knew and could do at various points on the scale for each subject.

Beginning in 1990, the NAGB initiated a process for defining *achievement levels* as regions along the overall reporting scale. Three levels were defined by minimum or cutoff scores: Basic, Proficient, and Advanced. Students who fail to reach the minimum score for the Basic achievement level are considered Below Basic. With these achievement levels, results can be reported in terms of percentages at or above a given level rather than as means on an arbitrary scale. Increases in the percentage of students who are Proficient (or have achieved at least basic mastery) are thought to be more meaningful for the general public and for policy setting than is an increase in the mean on the arbitrary scale.

Details of the achievement-level-setting process are well beyond the scope of this chapter. See NAGB (2000) for a recent discussion of achievement level standards. There has been some controversy about the process and the resulting achievement levels. Panels from the National Academy of Education (Shepard, 1993) and the National Research Council (Pellegrino, Jones et al., 1999) expressed concerns about the process and the resulting achievement level standards. The question is whether experts' judgments about particular students match the way NAEP standards would classify these students. For example, some students scoring 4 or 5 on an Advanced Placement Examination might not be classified as Advanced by the NAEP standards.

The process used to develop and adopt NAEP achievement level standards has evolved considerably over time. NAGB's current review procedures are designed to ensure a reasonable level of consistency across grades and subjects. Over time, the NAEP achievement levels will acquire their own meaning, whether or not they agree with other

conceptions of Basic, Proficient, or Advanced performance.

The NAEP achievement levels provide a useful benchmark for state efforts to define proficiency expectations. Estimates of the percentage of students at different achievement levels on state assessments can be compared with corresponding percentages from the NAEP state assessments. Discrepancies will doubtless lead to a political dialogue about the nature of the differences. There is often concern that state standards are too low, with the result that students are insufficiently challenged. Standards that are too high can be equally problematic, although this has been a less common concern. For example, where standards are too high, programs that may be working reasonably well might be abandoned in favor of riskier approaches that promise, but may not deliver, the inappropriately high levels of achievement that the standards require.

For local educators, standards-based reporting may not be important to instruction, at least until NAEP results for individual students or schools are included. Of greater use in shaping curriculum are the descriptions associated with each of the achievement levels. NAGB has established broad policy descriptions for each achievement level. As curriculum frameworks are developed, these policy descriptions are translated to statements about specific knowledge and skills associated with each of the achievement levels. These more detailed achievement level descriptions were originally developed by the standards-setting committees. With the 1996 science assessment, preliminary achievement level descriptions were added to the frameworks, with more explicit attention given to these descriptions in subsequent frameworks.

Sample Assessment Questions

Released NAEP questions and exercises reflect current thinking on how to assess accurately the knowledge and skills described in the content frameworks. A wealth of information about each item adds potential usefulness for educators. Anyone with Internet access can obtain this information from the NAEP questions section of the National Center for Education Statistics (NCES) website (http://nces.ed.gov/nationsreportcard/itmrls/). A list of questions is available for each subject and grade level. An advanced search option enables question selection by content area, ability, question type, or difficulty.

Clicking on a question in the list brings up the text of the question and provides options for viewing the following types of additional

information about it:

Performance data provides a graphic indicating the percentage of students answering the item correctly, or for open-ended questions with more than two score levels, the percentage of students at each score level.

Content classification indicates the content and ability categories the item represents and provides a description of these categories.

Scoring guide indicates the correct response option for multiple-choice questions. For open-ended questions that are hand-scored, the scoring rules or rubric is provided.

Student responses shows examples of actual student responses to the essay questions for different score levels.

More data indicates the percentage of students selecting each response option for multiple-choice questions or the percentage at each score level for open-ended questions. Response or score percentages are also disaggregated by gender, race, ethnicity, parents' education, type of school, region of the country, type of location, Title 1 participation, National School Lunch Program eligibility, and NAEP achievement level.

NAEP Question Example

The following example from the NAEP website illustrates the type of information that is available from the NAEP and how it might be used. Reading questions are organized around passages. One of the released passages for the fourth-grade assessment is titled "A Brick to Cuddle Up To." Students are asked to answer nine questions about this passage. The final question asks, "Does the author help you understand what colonial life was like? Use examples from the article to explain why or why not." Selecting this question on the NAEP website will display the full text of the passage, the text of the question, and five blank lines for student responses.

The *performance data* section for this item indicates that 20 percent of students provided responses that were judged as showing "evidence of full comprehension," 29 percent of the responses were judged as showing "evidence of partial or surface comprehension," and 51 percent were judged as showing "evidence of little or no

comprehension." We are also told that 0 percent skipped this item.

The *content classification* section tells us that the purpose of this question was "Reading to be informed" and the stance was "Demonstrating a critical stance." A paragraph describing each of these purposes is also provided. Several examples of question types are listed under the critical stance description. This question seems to match the type described as "How useful would this be for _____? Why?" although the question is not specifically tied to this type. A link to the reading framework is also provided in this section.

The *scoring guide* section provides descriptions of the basis for assigning responses to each of the three score levels. Under "Evidence of full comprehension," for example, it states:

> These responses provide an opinion about the author's abilities. In addition, they provide at least one supportive example from the text that demonstrates an objective consideration of the article and/or text-based critical judgment of the author's competence.

The *student responses* section provides examples of responses at each of the three scoring levels.

The *more data* section provides results separately for a wide variety of demographic groups. For example, 23 percent of students in the central region of the country got full credit for their responses while only 17 percent of students in the Southeast and West received full credit. This question appears to be relatively difficult for fourth graders in that among students at the Advanced achievement level, only 35 percent received full credit for their response. The question differentiates clearly between students at the Basic and at the Proficient levels. At the Basic level, 50 percent of the responses received the lowest score, and only 19 percent received full credit. At the Proficient level, 36 percent received the lowest score, and 29 percent received full credit.

Potential Uses of Sample Questions

One obvious use of released NAEP items is to embed them in classroom assessments. The supplemental information provided for each question will enable teachers to score responses, assess the types of questions (by content area or question format) that students can or cannot answer well, and compare classroom results to national outcomes.

Another potential use of the released questions is to provide concrete examples of the different areas of knowledge and skill covered in the content frameworks. This information may be useful to teachers

in designing instruction to cover these content areas. The questions might also form the basis of discussions with students about the skills they are expected to master.

Note, however, that some boundaries should be placed on teacher enthusiasm for using these questions. One limitation of the questions is that a small number of questions cannot provide a reliable indication of the consistency of a student's response across a range of stimuli and contexts. It is important not to value responses to a few released NAEP questions to the exclusion of information about students' performance over a substantial period of time.

A second limitation is that schools will vary in the extent to which their curriculum covers or is aligned with different areas of the NAEP content frameworks. An eighth grader's poor performance on algebra and functions questions may reflect the fact that he or she has not yet been taught many of the topics in this area covered by the NAEP assessment. In fact, analyses of student performance on NAEP items may reveal areas where local instruction could be expanded.

National Norms

At the heart of NAEP's design is nationally representative information about what students know and can do in different subjects and grades. Over time, we can see how much student achievement is improving and whether the percentage of students with significantly low levels of achievement is decreasing. We can also monitor trends in performance for specific subgroups of students, such as female achievement in mathematics or Hispanic achievement in reading. We can monitor trends at the state level. In this way, NAEP tells educators whether, as a whole, what we are doing is working.

A significant limitation of the normative information provided by NAEP is that there is currently no accepted way of obtaining NAEP scale scores or achievement level classifications for individual students. The Voluntary National Tests (VNT) proposed by President Clinton in 1997 were developed to assess fourth-grade reading and eighth-grade mathematics achievement relative to NAEP standards (Wise, Hauser, Mitchell, & Feuer, 1999). The tests were designed to be as consistent with NAEP in content and format as possible. Yet a panel commissioned by NAGB to examine methods for linking VNT scores to the NAEP scale expressed significant concerns about potential limitations (Cizek, Kenny, Kolen, & Van der Linden, 1999).

Another limitation of NAEP information is that it provides little diagnostic information about the specifics of what students do not understand or cannot do. NAEP was designed to maximize the accuracy in reporting overall achievement. Student-level assessments are generally more appropriate for diagnostic purposes. Two different committees of the National Academy of Sciences have discussed ways in which richer and more diagnostic information might be provided by NAEP and similar assessments (Pellegrino, Chudowsky, & Glaser, 2001; Pellegrino, Jones et al., 1999).

Until richer diagnostic information is available, educators can fall back on the wealth of normative information available on individual test items as described above. In many cases, released items can be found that demonstrate the specific knowledge and skills covered in particular lessons or curricular units. Comparison of individual student performance on such items with national norms can be useful diagnostic information that complements the summative information provided by overall NAEP results.

Potential Future Developments

NAEP is evolving. The currently proposed schedule for national and state assessments is shown in Table 2. A dramatic change, according to this plan, is that reading and mathematics will be assessed every year at the national or state level, although this assessment will be limited to grades four and eight. Another change is the introduction of a more comprehensive assessment with each introduction of a new or updated framework. New subjects, in particular a foreign language assessment for 12th graders, are also being added.

Pending federal legislation calls for using NAEP results to audit the achievement gains that states report based on their own assessments. The change to yearly assessment of reading and mathematics is designed to support this function, should it be enacted. Legislation establishing NAGB and allowing state reporting was passed by Congress in 1988 (i.e., the Hawkins-Stafford Elementary and Secondary School Improvements Amendments). The Improving America's Schools Act of 1994 further expands the role of NAEP. The No Child Left Behind Act mandated further participation in NAEP by the states and will lead to even greater attention to the state results.

NAEP as it exists today has great value for educators. As described here, the content frameworks, achievement level standards, and normative information are evidence of NAEP's value. Further, NAEP

Table 2. Assessments Scheduled from 1996 through 2012

	Years Assessed	
	National	State
Subject	**Year (Grades)**	**Grades 4 & 8**
Reading	1998 (4,8,12), 2000 (4) 2002 (4,8,12), 2003 (4,8) 2005 (4,8,12), 2007 (4,8) 2009 (4,8,12), 2011 (4,8)	1998 2002, 2003 2005, 2007 2009, 2011
Writing	1998 (4,8,12) 2002 (4,8,12) 2007 (8,12) 2011 (4,8,12)	1998* 2002 2007 2011
Mathematics	1996 (4,8,12) 2000 (4,8,12), 2003 (4,8) 2005 (4,8,12), 2007 (4,8) 2009 (4,8,12), 2011 (4,8)	1996 2000, 2003 2005, 2007 2009, 2011
Science	1996 (4,8,12) 2000 (4,8,12) 2005 (4,8,12) 2009 (4,8,12)	2000 2005 2009
U.S. History	2001 (4,8,12) 2010 (4,8,2)	
World History	2006 (12)	
Geography	2001 (4,8,12) 2010 (4,8,12)	
Economics	2006 (12)	
Civics	1998 (4,8,12) 2006 (4,8,12)	
Arts	1997 (8) 2008 (8)	
Foreign Language	2004 (12) 2012 (12)	
Long-Term Trend (Reading and Mathematics)	1996 (Ages 9, 13, 17) 2004 (Ages 9, 13, 17) 2008 (Ages 9, 13, 17) 2012 (Ages 9, 13, 17)	

* Assessed for grade 8 only.

has led to significant developments in the art of assessment, and released NAEP exercises provide useful examples and tools for educators seeking to design their own local assessments. It is ardently hoped that these aspects of NAEP's value will not be diminished as new functions and roles are added in future years.

References

Alexander, L., & James, H. T. (1987). *The nation's report card: Improving the assessment of student achievement.* Washington, DC: National Academy of Education.

Cizek, G. J., Kenny, P. A., Kolen, M. J., & Van der Linden, W. J. (1999). *Final report of the study group investigating the feasibility of linking scores on the proposed Voluntary National Tests and the National Assessment of Educational Progress.* Study commissioned by the National Assessment Governing Board, University of Toledo, OH.

Glaser, R., Linn, R., & Bohrnstedt, G. (1997). *Assessment in transition: Monitoring the nation's educational progress.* Stanford, CA: National Academy of Education.

Jones, L. V. (1996). A history of the National Assessment of Educational Progress and some questions about its future. *Educational Researcher, 25,* 15–22.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

NAGB [National Assessment Governing Board]. (2000). Student performance standards of the National Assessment of Educational Progress: Affirmation and improvements. Washington, DC: Author.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment* (Report of the Committee on the Foundations of Assessment, Board on Testing and Assessment, Commission on Behavioral and Social Sciences and Education, National Research Council). Washington, DC: National Academy Press.

Pellegrino, J. W., Jones, L. R., & Mitchell, K. J. (Eds.). (1999). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress* (Report of the Committee on the Evaluation of National and State Assessments of Educational Progress, Board on Testing and Assessment, Commission on Behavioral and Social Sciences and Education, National Research Council). Washington, DC: National Academy Press.

Shepard, L. (1993). *Setting performance standards for student achievement.* Stanford, CA: National Academy of Education.

Vinovskis, M. A. (1998). *Overseeing the nation's report card: The creation and evolution of the National Assessment Governing Board (NAGB).* Washington, DC: NAGB.

Wise, L. L., Hauser, R. M., Mitchell, K. J., & Feuer, M. J. (1999). *Evaluation of the Voluntary National Tests: Phase 1.* Washington, DC: National Academy Press.

Womer, F. B. (1970). *What is national assessment?* Ann Arbor, MI: NAEP.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☒ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").