



MARCH 2022

A Pragmatic Future for NAEP: Containing Costs and Updating Technologies

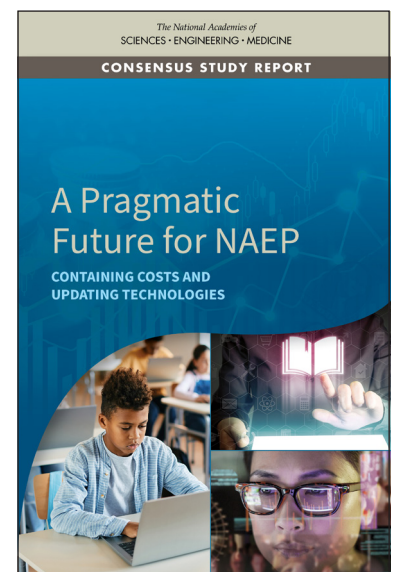
The National Assessment of Educational Progress (NAEP)—often called “The Nation’s Report Card”—is the largest nationally representative and continuing assessment of what students in public and private schools in the United States know and can do in various subjects and has provided policy makers and the public with invaluable information on U.S. students for more than 50 years. Given every two to four years, main NAEP assessments collect information on what 4th, 8th, and 12th graders in the United States know and can do in reading, mathematics, science, writing, and other academic subjects. A second program, “long-term trend” NAEP, measures the reading and mathematics achievement of 9-, 13-, and 17-year-olds.

Unique in the information it provides, NAEP is the nation’s only mechanism for tracking student achievement over time and comparing trends across states and districts for all students and important student groups (e.g., by race, sex, English learner status, disability status, family poverty status). However, the program has incurred substantially increased costs over the past two decades and currently costs about \$175.2 million per year. Moreover, its costs are rising, which has led to concerns about the program’s long-term viability.

Recent innovations in assessment technology that are increasingly used in state K–12 testing and other large-scale assessment programs have encouraged NAEP’s leaders to explore their feasibility for use with NAEP. To provide analysis and recommendations for the next phase of NAEP, the Institute of Education Sciences (IES) in the U.S. Department of Education requested that the National Academies of Sciences, Engineering, and Medicine form an expert panel to recommend innovations to improve the cost-effectiveness of NAEP while maintaining or improving its technical quality and the information it provides. The panel included members with expertise in psychometrics and educational measurement, artificial intelligence and machine learning, statistics, data science, and education and with experience with large-scale assessment.

UNDERSTANDING NAEP’S COSTS

The governance of NAEP is complex. The independent National Assessment Governing Board (NAGB) sets policy for NAEP, and the National Center for Education Statistics (NCES) administers the program. NAEP’s costs are spread between these two federal entities, each working with many contractors. The report recommends that NCES and



NAGB develop better descriptions of current spending on NAEP’s major components than are currently available and use them in making program decisions.

Management, planning, support, and oversight costs represent more than 28.7 percent of the total budget for NAEP, currently roughly \$50.3 million on average per year. This amount is very large, both in absolute terms and as a percentage of the overall NAEP budget. The report recommends that NAGB and NCES commission an independent audit of the program management and decision-making processes and costs, with a directive to review the program’s costs in detail and propose ways to streamline these processes.

STRUCTURAL CHANGES OF “MAIN NAEP” AND “LONG-TERM TREND NAEP”

Changing the Way Trends Are Monitored and Reported

The report recommends that NCES develop a proposal and detailed budget for modernization of long-term trend NAEP to include assessment frameworks and computer-based administration for 4th and 8th grade reading and mathematics. Congress, NAGB, and NCES should then consider the value of a modernized long-term trend NAEP in comparison with other program priorities.

Schedules for Updating Assessment Frameworks

Underlying each subject area tested by main NAEP are assessment frameworks, documents that lay out what is to be tested and how. To remain relevant, reflect contemporary content and methods, and to protect the ability to monitor trends over time, the report recommends that NAGB and NCES work both independently and collaboratively to achieve smaller and more frequent framework updates.

Integrating Complementary Assessments

Since its beginning, main NAEP has assessed subjects individually, but educational practice offers compelling arguments for combining assessments that test complementary subject matter. The report recommends that NAGB seriously consider the potential benefits of integrating the assessments for some complementary subjects, including reading and writing; science and technology and engineering literacy; and history, civics, economics, and geography.

ITEM DEVELOPMENT

Automated and Structured Item Development

Using a standardized process for creating test questions is one way of improving the efficiency of the test development process. Some testing programs use “automatic item generation,” which combines artificial intelligence/machine learning strategies and computer-based algorithms to generate test questions. These procedures work best when test questions are not complex, such as for simple, one-step, multiple-choice questions, but they are not likely to work well with NAEP’s more complex constructed-response and scenario-based questions. Instead of pursuing fully automatic item generation, the report recommends alternate strategies to enact a more structured approach to item development, such as with task models that depict the important features of items or the use of principled approaches to specify intended inferences and claims that items should assess.

Changing the Mix of Item Types

Traditionally, item types have been aligned with cognitive and content specifications, such that more complex item types have been used to assess more complex knowledge and skills. However, recent research calls this supposition into question – noting specifically that selected-response or simple constructed-response items can be effectively used to assess cognitively complex knowledge and skills. Since the simpler types of items tend to be less expensive to develop and score, the report recommends that NAGP commission an analysis of the costs and information provided by different item types.

MOVING TO LOCAL TEST ADMINISTRATION

The most expensive assessment component of NAEP’s budget is test administration. Representing 28.6 percent of NAEP’s budget, test administration presents one of the clearest opportunities for cost savings.

NCES has outlined a plan to transition to locally-based test administration in which school staff would serve as proctors and students would use the school's equipment for the NAEP assessment. The report recommends that NCES continue to develop and fully evaluate its plan for locally-based test administration. In addition, the report recommends that NCES develop specifications for the equipment, operating systems, and connectivity that are needed and that NCES collect the appropriate information about local devices and administration conditions on test day as well as explore statistical techniques to produce accurate estimates. Ultimately, full deployment of local administration may save substantially more than NCES currently estimates, which the report says should be reviewed.

INNOVATIONS IN TEST ADMINISTRATION

Testing Two Unrelated Subjects for Each Student

Currently, each student takes two blocks of test questions in one subject—for example, either reading or mathematics—and is given 30 minutes to respond to each block for a total of 60 minutes. Instead, the report recommends the program should increase testing time to 90 minutes to allow time to cover two different subjects and gather more information from each sampled student.

Reconsidering NAEP Sample Sizes

It may be possible for NAEP to report reliable information about student achievement and monitor performance gaps with smaller sample sizes. The report recommends that NCES analyze the tradeoffs between NAEP's sample sizes and statistical power for detecting educationally and politically meaningful differences in performance. While some comparisons require larger samples, such as for students grouped by race/ethnicity and socioeconomic status, others do not. Careful determination of the sample sizes needed to support policy-relevant inferences may suggest smaller sample sizes and cost reductions.

ADAPTIVE TESTING

Adaptive testing uses computer-based algorithms to update the estimate of the student's proficiency level during the test and uses the estimate to select questions that are neither too simple nor too difficult for the student to answer. The adjustments can be made after each question or after groups of questions, with the latter method being more suited to NAEP, which often groups multiple questions related to a single passage or stimulus. However, because sets of items typically include a mix of easy and hard items, the panel is not optimistic that this method will result in substantial efficiencies across the full population. The report recommends that NCES should not pursue adaptive testing for NAEP as a way of saving costs but should continue to investigate its potential to improve achievement estimates and the test-taking experiences for low-performing students.

AUTOMATED ITEM SCORING

The report recommends that NAEP continue its work to automate the scoring of certain item types on the 4th and 8th grade reading and mathematics assessments. Automated scoring has the potential to reduce costs, improve the speed of reporting, increase the information provided about open-ended responses, increase the consistency of scoring over time, and provide information about fairness. Automated scoring is likely to be cost-effective for the NAEP assessments with large state-level samples, which could reduce costs by about 0.7 percent of NAEP's budget.

ANALYSIS AND REPORTING

Arguably, all of NAEP's impact is mediated through analysis and reporting, which include not only score reports and related data and analyses, but also the program's frameworks, innovative example items, advanced psychometrics, and other assessment practices. NAEP score reports have regularly provided clear, high-level overviews of NAEP results. Reports that go a step further and analyze relationships between NAEP data and data from other sources are consistently among the most popular reports produced with NAEP data.

The report recommends that NCES devote a greater percentage of its budget for innovative analysis and reporting to increase the use and understanding of NAEP's data. This should include finding ways to make the raw data available more quickly to researchers, improving the usability and sophistication of the NAEP Data Explorer, making process data more easily accessible, and expanding the availability and use of important contextual variables.

VISION FOR A TECHNOLOGICAL AND RESEARCH INFRASTRUCTURE

The platform and technology currently used to deliver NAEP assessments on computers is almost a decade old and is based in a customized application that requires dedicated tablet computers, dedicated internet routers, and contractor staff at every school site in which NAEP is administered. Developing an up-to-the-date platform is necessary to administer NAEP on local computers, as well as to take advantage of other promising innovations in the years ahead.

The report recommends that NCES systematically evaluate the software built by vendors or available in open-source libraries for its potential to meet the requirements of the different components of its new technology platform. Those components should be custom built only if rigorous analysis shows large net benefits from doing so and the decision should be made on a component basis, not as a single decision to build or buy all components. The expertise needed to carry out these analyses is substantially different than the expertise in psychometrics and statistics that is prevalent among NCES staff members. NCES will need to ensure that there is adequate internal and external expertise in software development to support and oversee this work.

The report also recommends that NCES increase the visibility and coherence of NAEP's research activities to help NAEP's stakeholders understand the innovations the program is investigating and the lessons it is learning. The NAEP research program should have an identifiable budget and program of activities.

Opportunities for the National Assessment of Educational Progress in an Age of AI and Pervasive Computation: A Pragmatic Vision

KAREN J. MITCHELL (*Chair*)

Association of American Medical Colleges (retired)

ISSAC I. BEJAR

Educational Testing Service (retired)

SEAN PATRICK (JACK) BUCKLEY

Roblox, New York, NY

BRIAN GONG

Center for Assessment, Dover, NH

ANDREW D. HO

Harvard Graduate School of Education

STEPHEN LAZER

Questar Assessment Incorporated, Cape May, NJ

SUSAN M. LOTTRIDGE

Cambium Assessment, Inc., Harrisonburg, VA

RICHARD M. LUECHT

School of Education, University of North Carolina at Greensboro

ROCHELLE S. MICHEL

Curriculum Associates, Lawrenceville, NJ

SCOTT NORTON

Council of Chief State School Officers, Baton Rouge, LA

JOHN WHITMER

Federation of American Scientists, Davis, CA

STUART W. ELLIOTT

Study Director

JUDITH KOENIG

Senior Program Officer

ANTHONY MANN

Program Associate

For More Information . . . This Consensus Study Report Highlights was prepared by the Committee on National Statistics based on the Consensus Study Report *A Pragmatic Future for NAEP: Containing Costs and Updating Technologies* (2022). The study was sponsored by the U.S. Department of Education. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project. Copies of the Consensus Study Report are available from the National Academies Press, (800) 624-6242 or <https://www.nap.edu/catalog/26427>.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.

www.nationalacademies.org