

NAEP Budget Workshop

CLOSED SESSION

On Friday afternoon, August 5, during closed session Board members will participate in a budget workshop. The National Center for Education Statistics (NCES) will present information on how NAEP Alliance contracts were planned and awarded for the 2019 cycle. NCES will also provide confidential information about NAEP costs, potential cost savings, and future budget planning.

NCES has produced two white papers (Attachments A and B) as background information on the NAEP Alliance contracts and NAEP costs. These white papers are public information and do not disclose proprietary or confidential details about NAEP.

The session will begin with an NCES presentation, which will be followed by small group breakouts among members. The session will conclude with key takeaways from the breakout discussions and the full Board identifying next steps and topics for future meeting agendas related to the NAEP budget.

Overview of the NAEP Contracts Alliance and How the Contracts are Competed

The “NAEP Alliance” is an interdependent group of contracts NCES uses to administer the National Assessment of Educational Progress (NAEP). The work is carried out through several contracts rather than a single one, because the scope of work is interrelated, as each component is closely tied with others and has dependencies that cannot be separated without impact on the overall scope. This paper describes the current NAEP Alliance and provides a brief history of how the Alliance concept has developed over time as the Contracts and Acquisition Management Office (CAM) and NCES have worked to encourage competition and broaden the range of vendors supporting NAEP, while maintaining high quality work and controlling costs.

The Alliance contracts are focused on work that is so interdependent that failures in cooperation would produce substantial risk to the success of NAEP. Development and implementation of any large-scale assessment involves a series of highly technical steps with tight timelines and required handoffs from team to team (e.g., assessment items are written by subject matter experts, but with critical handoffs with psychometricians and the digital assessment platform programmers, among others). The Alliance covers about three quarters of the work on NAEP in terms of dollars and the remaining quarter is undertaken through several other, independent contracts.

For each contract cycle, NCES and CAM consider how to structure the contracts to encourage competition, increase the range of vendors working on NAEP (including small businesses), and maximize quality and cost efficiency. For the last few cycles an Alliance contract structure has been determined to be the best option, although the structure itself has changed.

The Alliance competition is a series of individual, contract-by-contract competitions, each scored independently. NAEP Alliance proposals are not scored solely as individual bids; a team proposal is also evaluated. The winning bid is the team with the best set of individual *and* team proposals. All proposals must be of high technical quality and acceptable cost.

Competition of Alliance contracts is centered on individual competitions for each separate contract. Bidders must be part of a team because of the interdependencies among the various workstreams, but their technical and cost proposals are reviewed and scored solely within the single contract being competed. Separate evaluator panels review the bids for each contract. Each team must also submit a “team volume” describing the team’s approach to working together which is also scored. In the end, the winning bidders are those whose “offer as a team conforms to the solicitation and will be most advantageous to the Government, price/cost and other factors considered (for detail on the 2019 solicitation, see solicitation FORMSF33-91990018R0018 at the SAM.gov [solicitation notice](#)).” The final

Two of the four Alliance competitions have received multiple bids – on par with other NCES procurements.

technical and cost details are negotiated independently for each contract. In this way, NAEP obtains contractors that have both the capability to do the work assigned to them and the ability to work together as a team toward a common goal. Each requirement—individual contractor capability and the demonstrated ability to work as a team—is a necessary requirement but not sufficient on its own. Both requirements must be satisfied before a bidder is eligible for award of a contract.

Over the four contract cycles in which CAM and NCES have employed an alliance structure, the government has received multiple bids twice—or 50 percent of the time. This is on par with the percentage of procurements receiving multiple bids across NCES.

NCES has also kept costs down over the four cycles. In real terms (constant 2021 dollars), NAEP’s program costs in 2020 were the same as program costs in 2002¹, despite a dramatic increase in the

NAEP costs have not increased in real terms in 20 years.

scope of the program, including the transition to a digitally based assessment and nearly tripling the size of NAEP’s Trial Urban District Assessment (TUDA) of select large urban school districts. Still, CAM and NCES are always seeking ways to increase competition and cost efficiency.

Team Structure for NAEP

The 2019 – 2024 NAEP Alliance consists of nine contracts, as described in table 1. As discussed above, the Alliance contracts are intended only to cover activities that are highly interdependent and have a significant risk for failure if

contractors do not cooperate effectively. The current Alliance contracts range from support for NCES in planning and coordinating NAEP work to design and development to the administration of the assessments, data collection, data analysis and reporting results. Other work is contracted through several independent, non-Alliance contracts, including contracts for communications and outreach, logistics, secondary analysis and reporting, the funding of the NAEP state coordinators, and some quality assurance work.

The current Alliance includes five prime contractors: ETS, Huntington Ingalls Industries, Management Strategies, Pearson, and Westat. Management Strategies is a registered small business.

CAM and NCES have employed a team contract structure since 2002. Similar team structures are

Table 1. Current NAEP Alliance Contracts

Contract Title	Scope of Work	Prime Contractor
Planning and Coordination	Coordinating communication across contractors and NCES; strategic planning	ETS
Program Support Management	Scheduling, quality control, risk management	Management Strategies
Design, Analysis and Reporting	Design of studies; analysis; specification of data needed to meet reporting goals; reports	ETS
Item Development	Cognitive items, scoring rubrics and survey questions; assistance in training of scorers; cognitive interviews/small-scale pilots; translation	ETS
Sampling and Data Collection	Sampling; sampling weights; administration of assessments; data collection; transcript studies	Westat
Support and Service Center	Support, training, and resources to state and TUDA coordinators	Westat
Materials, Distribution, Processing and Scoring	Assessment and auxiliary materials; distribution of assessment booklets and materials; scoring training materials; scoring	Pearson
Web/Technology Development, Operations, and Maintenance	Internet-related applications and services; web and other computer-based products and services	Huntington Ingalls Industries
NAEP Platform Development	Assessment delivery platform	ETS

¹ In 2021 dollars, the NAEP budget in 2002 was \$162M, compared with \$160M in 2020.

employed by the Department of Defense, the international largescale assessments such as the Program for International Student Assessment (PISA) and the Program for the International Assessment of Adult Competencies (PIAAC), among others.

Team structures like those employed by NCES and the Department of Defense have several key elements that reduce risk of project failure and make them in many cases in the best interest of the government:

- Contracts include stipulations and incentives for successful cooperation. For example, team award fees (and penalties) for individual NAEP contracts are tied to the program’s success in reporting assessment results within 6 months of the completion of data collection. Reporting results this quickly can only be accomplished through successful cooperation across contractors.
- Bidders form teams before submitting proposals. The decision on which partners come together as a team is a matter of bidder choice. This helps ensure overall quality of the team and their ability to work together (bidders “vet” each other as they form a team), improves the quality of proposals, as bidders work concretely through the many handoffs required of the contracts, and provides an opportunity for bidders to develop as a team as they write their proposals.
- The proposals for each contract are rated and negotiated independently; no contract is signed by the government unless it is technically sound and reasonably priced—regardless of the technical quality and price of the team members’ proposals for other contracts. Weak links are not acceptable.

The team model requires NCES to appoint a contracting officer’s representatives to monitor each contract (i.e. component of the work), which allows the government to learn over time how the contracts can be restructured to be more efficient and what work can be extracted from the Alliance and competed independently to develop small business engagement, grow the number of vendors supporting NAEP, and ultimately increase competition.

Employing multiple contracts to implement NAEP also supports the program in building the number of vendors, including small businesses, involved in NAEP. For example, CAM and NCES have been able to carve out work in the Alliance that has been won by small businesses as prime contractors. Also, using multiple contracts allows the program to support many vendors as subcontractors (including small businesses) without having too many subcontractors per contract to manage effectively.

This has allowed CAM and NCES to greatly expand the range of vendors with experience on NAEP. By 2019, the Alliance model involved 51 contractors in total, including 29 small businesses. Subcontracted work ranges from printing and logistics to translation to development of sophisticated web-based analytic tools.

51 vendors are part of the current NAEP Alliance.

Efforts to Increase Competition in 2019

Although competition for the NAEP Alliance is similar to other NCES procurements in terms of numbers of bids received, CAM and NCES worked hard in 2019 to increase competition. In addition to market research undertaken by NCES technical staff to search for vendors with capabilities in the fields covered by NAEP, CAM convened a large Vendor Communication Conference in an effort to generate interest in the procurement, educate potential vendors on the planned scope and structure of the contracts, and

gather input from potential vendors to improve the solicitation and gain more and higher quality proposals.

CAM posted draft performance work statements written by NCES staff 9 months prior to the release of the solicitation and a month before the Vendor Communication Conference. The Vendor Communication Conference took place 8 months prior to release of the solicitation. The long lead time prior to the final solicitation was intended to allow inexperienced vendors to learn more about NAEP and prepare for a potential proposal, as well as to gain their insights on how to better compete the contracts. CAM and NCES responded to vendors' questions from the conference and used input obtained from the vendors to improve the final performance work statements and other solicitation documents.

Looking Forward

In preparation for the next NAEP contract cycle in 2024, NCES is again working with CAM to consider the contract structure and other elements of the planned solicitation. In addition, CAM will soon post a draft statement of objectives for a cost structure review of NAEP's contracts and then award a contract for the work in the fall. The cost structure review is a comprehensive review of NAEP's structures and processes from assessment design through reporting to identify inefficiencies and provide NCES recommendations for alternative ways to implement NAEP. NCES will use the results of this work in preparing for the 2024 NAEP contracts procurement. Ultimately, we must find a model that retains the benefits of the Alliance structure, keeps costs reasonable, keeps risk of program failure low (including risks to NAEP's long trend lines on student performance), and allows for growth and innovation.

Understanding NAEP Operational Costs

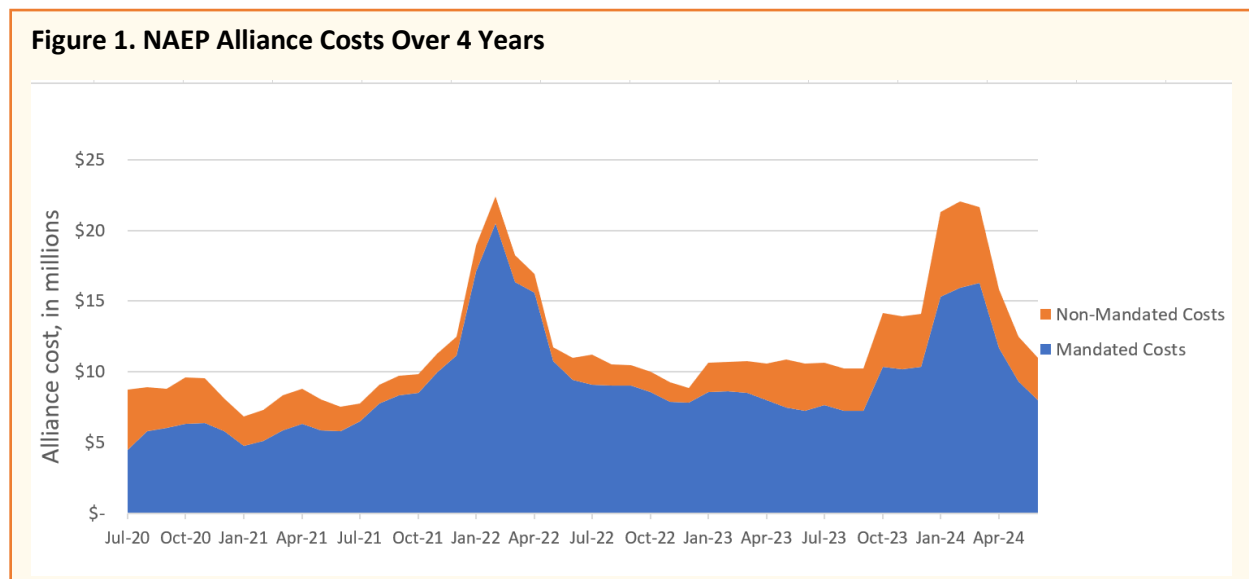
NAEP operations, from the development of assessments through data collection through the reporting of results, are designed to ensure successful implementation of the [NAEP assessment schedule](#) with cost efficiency and the capability to accommodate changes in the schedule. The schedule is set by the National Assessment Governing Board. The centerpiece of the schedule and NAEP's operations is the large, congressionally mandated biennial combination of assessments in mathematics and reading at grades 4 and 8 for the nation, states, and select large urban districts. These assessments form the core of NAEP operations and costs. NCES has built an infrastructure for high quality, cost-efficient implementation of these mandated assessments, as well as other assessments¹ that appear regularly on the NAEP assessment schedule.

NAEP's Cost Cycle

Implementation of the mandatory state-level assessments in mathematics and reading creates a relatively regular two-year cost cycle. Figure 1 shows two of these cycles for the NAEP Alliance contracts that perform the bulk of this work. The figure shows separately costs for the mandated mathematics and reading assessments and for all other assessments and activities. There are two key findings that stand out from the figure.

The Congressionally mandated assessments in mathematics and reading, grades 4 and 8, make up 78 percent of total NAEP Alliance costs.

First, the mandated assessments make up 78 percent of total NAEP Alliance costs over the four-year period. This reflects in part the frequency of the mandated assessments and their size in terms of numbers of students and schools sampled. Mathematics and reading at grades 4 and 8 are administered



¹Other regular assessments include grade 12 mathematics and reading assessments, science assessments at grades 4, 8, and 12, grade 8 civics and U.S. history assessments, and the long-term trend assessments in reading and mathematics at ages 9, 13, and 17. These are all typically administered at the national level approximately every 4 years.

every two years and to samples designed to represent the nation, the states, DC, Puerto Rico, the Department of Defense Education Activity (DODEA), and more than 20 Tribal Urban District Assessment (TUDA) districts. The non-mandated assessments typically happen four or more years apart and usually only to represent the nation. However, the high relative cost of the mandated assessments reflects also the fixed costs of systems put in place to administer the mandated assessments but used for the efficient implementation of other assessments, as well.

Second, costs are higher in the years in which the mandated assessments are administered, but the other years still have significant costs. For the period covered by figure 1, the average Alliance costs in years with the mandated assessments is \$170 million, compared with \$110 million in the other years. Major costs in the mandated assessment years include data collection, scoring, analysis, and reporting. The work that occurs in the other years includes sampling, state and school outreach, transcript studies, and development of the assessments and surveys, including R&D, item development (including item writing and pretesting), and trend maintenance of existing items.

Other Major Cost Factors

Although the mandated assessments make up the lion's share of NAEP costs, other factors play a major role. These factors include implementation of new assessment frameworks and the number and timing of other, non-mandated assessments. Implementation of a new or updated framework typically takes 5 years of development before the operational assessment. New or updated frameworks (such as those to be assessed in reading and mathematics in 2026) require thorough review and understanding of the framework and item specifications. After adoption, extensive implementation plans are written to align new development to the overall assessment design, including the number and types of items to be developed, the format of the assessment, new digital assessment tools that will be made available to students, and various reporting goals. Research and investigations are often required to validate new item types or to examine how best to operationalize new and expanded constructs. Content and platform development then begin to meet these needs and include determining the alignment of existing item pools to the new framework, writing new items, various stages of content, editorial, accessibility, translatability, and functionality reviews, including Standing Committee, sensitivity and fairness, stakeholder (e.g., States and Districts) and the Governing Board; and item pretesting with students. The new content and platform are then pilot-tested two years prior to the operational assessment to collect data that are used to select items/tasks and assemble blocks that will meet the goals of the framework.

Other than the mandated assessments, **major cost factors** include **implementation of new frameworks** and whether additional assessments require **standalone administration** or are extensions of field work already planned.

NCES has ongoing work to develop and implement all assessments on the NAEP assessment schedule, including several that appear regularly on the schedule, such as mathematics and reading at grade 12, science at grades 4, 8, and 12, civics and U.S. history at grade 8, and long-term trend assessments in mathematics and reading at ages 9, 13, and 17.

The Governing Board and NCES work together to integrate these assessments into the schedule to optimize cost-efficiency and maximize attention on the results of these assessments. Much of the infrastructure of NAEP is needed to support the mandated assessments. For example, a representative sample of 8th grade schools is required regardless of whether NAEP is assessing reading and

mathematics or reading, mathematics, U.S. history, and civics. As such, it is often more efficient to add a national-level assessment in assessment years in which mandated assessments are already being administered. That said, a sample of 12th grade schools is a different sample and would need to be added separately regardless of whether there are also assessments in grades 4 and 8.

Cost-Saving Modernizations

Research and development are underway at NCES to pursue modernizations that promise to significantly reduce operational costs in the latter half of the decade. These include

- (1) **Device agnostic NAEP:** administration of NAEP on any device that meets a set of technical standards, rather than exclusively on NAEP-provided Microsoft Surface Pro tablets;
- (2) **Reduced contact:** reduction of field staff sent to schools, potentially placing some burden for administration on school staff;
- (3) **Automated scoring:** introduction of artificial intelligence (AI) tools to expand the use of automation in the scoring of constructed response items.

Device Agnostic

In 2022, NCES completed a proof-of-concept study on the next generation of the eNAEP digital assessment platform that will support administration of NAEP on devices other than the Surface Pros

Device agnostic NAEP is projected to save \$27 million in one-time costs in the transition to less expensive devices. As NAEP shifts to school-based equipment, savings will grow.

currently used. Upon successful completion of a field test of the new digital platform in 2023, NCES plans to administer the 2024 assessments using a combination of Surface Pros and less expensive NAEP-provided Chromebooks and to administer the 2026 assessments using a combination of NAEP-provided Chromebooks and school-based equipment.

The first savings will occur in 2024 when NCES begins replacing the Surface Pros with Chromebooks. By the time the fleet of Surface Pros is replaced in 2026, the program will have saved \$27 million, because of the difference in procurement costs. Additional savings will accrue in 2026 (\$1 million) and 2028 (\$2 million) from using increasingly more school-based equipment and fewer NAEP-provided devices. When it is time to replace the new fleet of Chromebooks (post-2028), the program should see additional savings because fewer devices will be procured.

Reduced Contact

The 2023 eNAEP digital platform field test will include testing the administration of NAEP with fewer field staff per school. Development and integration of eNAEP and the Assessment Management System that supports the administration of NAEP have dramatically streamlined processes for field staff in schools as they administer NAEP. As a result, NCES expects to reduce the number of field staff per school by one person, resulting in 25 percent fewer field staff. The reduction in field staff is projected to save NAEP \$4 million in 2024. This form of reduced contact is called Reduced Contact 1 (RC1).

Reduced Contact is projected to **save \$4 million** in 2024 **without adding to school burden**. In 2028, with some **burden shifted to schools, savings will reach \$7M**.

NCES is developing a Reduced Contact 2 (RC2) format that will reduce further the number of field staff required per school but will require the shifting of some administration duties to school staff. In this format, one field staffer, an Assessment Coordinator, will be sent to each participating school and one or

more school staff members will take up the responsibilities of Assessment Administrators. This format will require increased training and help desk support and therefore the savings per-field-staff-removed are not as steep. Under the assumption that we can implement RC2 in 90 percent of schools by 2028, NCES projects savings of \$7 million in 2028 and each administration going forward.

Automated Scoring

NCES fielded a successful challenge in 2022 in which competitors were asked to score, using AI tools, 22 constructed response items taken from the 2017 reading assessments at grades 4 and 8. Challengers' AI-derived scores were compared for accuracy with the scores assigned to responses by human scorers. Based on the success of the challenge, NCES is moving forward with a study to re-score 2022 reading constructed response items using AI methods. If that study is successful, NCES plans to include automated scoring alongside human scoring as a bridge study in 2024 and then incorporate automated scoring as the primary scoring method (some human scoring is expected always to be required) in 2026 and beyond. Work is proceeding also on introducing AI to expand automated scoring in NAEP mathematics.

Automated Scoring is projected to **save about \$2 million** per administration.

Savings from automated scoring are expected to begin in 2026 (\$1.7M) and grow in 2028 (\$2.2M) as automated scoring expands in mathematics.

Modernizations Summary

Table 1 lists the modernization efforts along with key assumptions on which projected savings are based. The biggest savings over the current administration model are expected to result from the move to device agnostic administration of NAEP. This includes \$27 million from switching the NAEP device fleet over to less expensive devices, as well as additional savings if fewer NAEP-provided devices are sent into the field. When it is time to replace the new fleet of NAEP-provided devices, the program

should attain additional savings by replacing fewer and shrinking the size of the fleet.

The savings from reduced contact depends on the extent to which the program is

Table 1. Projected costs and savings for ongoing NAEP innovations

Innovation	Assumptions	Savings		
		2024	2026	2028
Device agnostic	Chromebooks phased in	\$27M		
- Cheaper NAEP devices - Use of school equip.	90% of schools using school equipment in 2028	--	\$1M	\$2M
Reduced contact	90% of schools using school staff to administer in 2028	\$4M	\$5M	\$7M
Automated scoring	80% of reading, 70% math autoscored by 2028	--	\$1.7M	\$2.2M

willing to shift burden to schools. Without shifting any burden to schools, the program can save \$4 million per administration. If contact in schools is reduced to one field staffer per school in 90 percent of schools, then the program can almost double the savings. However, this depends on the extent to which schools take on administration responsibilities currently handled by NAEP field staff. Automated scoring is expected to save \$2 million per administration by 2028.

Looking Forward

The cost-saving modernizations in development for NAEP represent major structural changes in program operations. Some savings can be obtained directly from technological advances, including the move to more generic, less expensive devices, streamlined administration that allows for fewer field staff, and the use of artificial intelligence to extend automated scoring. Other savings, however, can only be obtained through changes in NAEP's relationship with schools. This includes greater use of school-based devices and Internet access, rather than NAEP-provided devices and Internet, as well as potentially greater reliance on school staff for the administration of NAEP. These advances will require thoughtful deliberation about the program's relationship with schools, as well as outreach and discussion with schools, districts, and states.