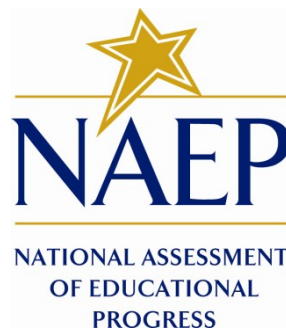


*NATIONAL CENTER FOR EDUCATION STATISTICS
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS*

NAEP Assessments for 2017-2019

*SUPPORTING STATEMENT
PART A*

*OMB# 1850-NEW v.1
(previous OMB# 1850-0790 v.43)*



November 2015
Revised August 2016

Table of Contents

A.1. Circumstances Making the Collection of Information Necessary.....	3
A.1.a. Purpose of Submission.....	3
A.1.b. Legislative Authorization.....	4
A.1.c. Overview of NAEP Assessments.....	4
A.1.c.1. NAEP Frameworks.....	4
A.1.c.2. Cognitive Item Development.....	5
A.1.c.3. Survey Items.....	5
A.1.c.4. Inclusion in NAEP.....	7
A.1.c.5. Transition to Digitally Based Assessments (DBAs).....	8
A.1.c.6. Assessment Types.....	11
A.1.d. Overview of 2017-2019 NAEP Assessments.....	12

A.2. How, by Whom, and for What Purpose the Data Will Be Used.....	17
A.3. Improved Use of Technology.....	18
A.4. Efforts to Identify Duplication.....	20
A.5. Burden on Small Businesses or Other Small Entities.....	20
A.6. Consequences of Collecting Information Less Frequently.....	21
A.7. Consistency with 5 CFR 1320.5.....	21
A.8. Consultations Outside the Agency.....	21
A.9. Payments or Gifts to Respondents.....	22
A.10. Assurance of Confidentiality.....	23
A.11. Sensitive Questions.....	26
A.12. Estimation of Respondent Reporting Burden (2017–2019).....	26
A.13. Cost to Respondents.....	32
A.14. Estimates of Cost to the Federal Government.....	33
A.15. Reasons for Changes in Burden (from last Clearance submittal).....	33
A.16. Time Schedule for Data Collection and Publications.....	33
A.17. Approval for Not Displaying OMB Approval Expiration Date.....	34
A.18. Exceptions to Certification Statement.....	34

Appendixes

- Appendix A External Advisory Committees
- Appendix B Sample Data Security Agreement
- Appendix C 2017 Sampling Memo
- Appendix D Communications and Recruitment Materials
- Appendix E Assessment Feedback Forms
- Appendix F NAEP Survey Questionnaires
- Appendix G NAEP 2011 Sampling Design
- Appendix H NAEP 2011 Weighting Procedures

A.1. Circumstances Making the Collection of Information Necessary

A.1.a. Purpose of Submission

The National Assessment of Educational Progress (NAEP) is a federally authorized survey of student achievement at grades 4, 8, and 12 in various subject areas, such as mathematics, reading, writing, science, U.S. history, civics, geography, economics, technology and engineering literacy (TEL), and the arts.

NAEP is conducted by the National Center for Education Statistics (NCES) in the Institute of Education Sciences of the U.S. Department of Education. As such, NCES is responsible for designing and executing the assessment, including designing the assessment procedures and methodology, developing the assessment content, selecting the final assessment content, sampling schools and students, recruiting schools, administering the assessment, scoring student responses, determining the analysis procedures, analyzing the data, and reporting the results¹.

The National Assessment Governing Board (henceforth referred to as the Governing Board), appointed by the Secretary of Education but independent of the Department, is a bipartisan group whose members include governors, state legislators, local and state school officials, educators, business representatives, and members of the general public. The Governing Board sets policy for NAEP and is responsible for developing the framework and test specifications that serve as the blueprint for the assessments.

The NAEP assessments contain diverse items such as “cognitive” assessment items, which measure what students know and can do in an academic subject, and “survey” or “non-cognitive” items, which gather factual information such as demographic variables, as well as construct-related information, such as courses taken. The survey portion includes a collection of data from students, teachers, and school administrators.

Since NAEP assessments are administered uniformly using the same sets of test booklets across the nation, NAEP results serve as a common metric for all states and select urban districts. The assessment stays essentially the same from year to year, with only carefully documented changes. This permits NAEP to provide a clear picture of student academic progress over time.

NAEP consists of two assessment programs: the NAEP long-term trend (LTT) assessment and the main NAEP assessment. The LTT assessments are given at the national level only, and are administered to students at ages 9, 13, and 17 in a manner that is very different from that used for the main NAEP assessments. LTT reports mathematics and reading results that present trend since the 1970s. Within the timeframe covered under this submission, only main NAEP assessments will be administered.

NAEP provides results on subject-matter achievement, instructional experiences, and school environment for populations of students (e.g., all fourth-graders) and groups within those populations (e.g., female students, Hispanic students). NAEP does not provide scores for individual students or schools. The main NAEP assessments report current achievement levels and trends in student achievement at grades 4, 8, and 12 for the nation and, for certain assessments (e.g. reading and mathematics), states and select urban districts. The Trial Urban District Assessment (TUDA) is a special project developed to determine the feasibility of reporting district-level results for large urban districts. Currently, the following districts participate in the TUDA program: Albuquerque, Atlanta, Austin, Baltimore City, Boston, Charlotte, Chicago, Clark County (NV), Cleveland, Dallas, Denver, Detroit, District of Columbia (DCPS), Duval County (FL), Fort Worth, Fresno, Guilford County (NC), Hillsborough County (FL), Houston, Jefferson County (KY), Los Angeles, Miami-Dade, Milwaukee, New York City, Philadelphia, San Diego, and Shelby County (TN).

¹ The role of NCES, led by the Commissioner for Education Statistics, is defined in 20 U.S. Code Section 9622 (<https://www.law.cornell.edu/uscode/text/20/9622>) and OMB Statistical Policy Directives No. 1 and 4 (https://www.whitehouse.gov/omb/inforeg_statpolicy).

This submission requests OMB's approval for the following NAEP 2017-2019 assessments: operational, pilot, and special studies.

A.1.b. Legislative Authorization

In the current legislation that reauthorized NAEP, the National Assessment of Educational Progress Authorization Act (20 U.S. Code Section 9622), Congress mandates the collection of national education survey data through a national assessment program:

1. ESTABLISHMENT- The Commissioner for Education Statistics shall, with the advice of the Assessment Board established under section 302, carry out, through grants, contracts, or cooperative agreements with one or more qualified organizations, or consortia thereof, a National Assessment of Educational Progress, which collectively refers to a national assessment, State assessments, and a long-term trend assessment in reading and mathematics.
2. PURPOSE; STATE ASSESSMENTS-
 - (1) PURPOSE- The purpose of this section is to provide, in a timely manner, a fair and accurate measurement of student academic achievement and reporting of trends in such achievement in reading, mathematics, and other subject matter as specified in this section.

The National Assessment of Educational Progress Authorization Act also requires the assessment to collect data on specified student groups and characteristics, including information organized by race/ethnicity, gender, socio-economic status, disability, and English language learners. This allows for the fair and accurate presentation of achievement data and permits the collection of background, non-cognitive, or descriptive information that is related to academic achievement and aids in the fair reporting of results. The intent of the law is to provide representative sample data on student achievement for the nation, the states, and a variety of populations of students, and to monitor progress over time.

The statute and regulation mandating or authorizing the collection of this information can be found at <https://www.law.cornell.edu/uscode/text/20/9622>.

A.1.c. Overview of NAEP Assessments

This section provides a broad overview of NAEP assessments, including information on the assessment frameworks, the cognitive and survey items, inclusion policies, the transition to digitally based assessments (DBAs), and the assessment types.

A.1.c.1. NAEP Frameworks

NAEP assessments follow subject-area frameworks developed by the Governing Board and use the latest advances in assessment methodology. Frameworks capture a range of subject-specific content and thinking skills needed by students in order to deal with the complex issues they encounter inside and outside their classrooms. The NAEP frameworks are determined through a framework development process that ensures they are appropriate for current educational requirements. Because the assessments must remain flexible to mirror changes in educational objectives and curricula, the frameworks must be forward-looking and responsive, balancing current teaching practices with research findings.

NAEP frameworks can serve as guidelines for planning assessments or revising curricula. These frameworks also can provide information on skills appropriate to grades 4, 8, and 12 and can be models for measuring these skills in innovative ways. The subject-area frameworks evolve to match instructional practices.

Developing a framework generally involves the following steps:

- widespread participation and reviews by educators and state education officials;
- reviews by steering committees whose members represent policymakers, practitioners, and members of the general public;
- involvement of subject supervisors from education agencies;
- public hearings; and
- reviews by scholars in the field, by NCES staff, and by a policy advisory panel.

The frameworks can be found at <https://www.nagb.org/publications/frameworks.html>.

A.1.c.2. Cognitive Item Development

As part of the item development process, NCES calls on many constituents to guide the process and review the assessment. Item development is guided by a multi-year design plan, which is guided by the framework and establishes the design principles, priorities, schedules, and reporting goals for each subject. Based on this plan, the NAEP contractor creates a development plan outlining the item inventory and objectives for new items and then begins the development process by developing more items than are needed. This item pool is then subjected to:

- internal contractor review with content experts, teachers, and experts on political sensitivity and bias;
- playtesting, tryouts, or cognitive interviews with small groups of students for select items (particularly those that have new item types, formats, or challenging content); and,
- refinement of items and scoring rubrics under NCES guidance.

Next, a standing committee of content experts, state and local education agency representatives, teachers, and representatives of professional associations reviews the items. The standing committee considers:

- the appropriateness of the items for the particular grade;
- the representative nature of the item set;
- the compatibility of the items with the framework and test specifications; and
- the quality of items and scoring rubrics.

For state-level assessments, this may be followed by a state item review where further feedback is received. Items are then revised and submitted to NCES and the Governing Board Assessment Development Committee for approval prior to pilot testing.

The pilot test is used to finalize the testing instrument. Items may be dropped from consideration or move forward to the operational assessment. The item set is once again subjected to review by the standing committee, the Governing Board, and NCES following generally the same procedure described above. A final set of test items is then assembled for NCES and the Governing Board's review and approval.

After the operational assessment, items are once again examined. In rare cases where item statistics indicate remaining problems, the item may be dropped from the assessment. The remaining items are secured for reuse in future assessments, with a subset of those items publicly released.

A.1.c.3. Survey Items

In addition to assessing subject-area achievement, NAEP collects information that serves to fulfill the reporting requirements of the federal legislation and to provide context for the reporting of student performance. The legislation requires that, whenever feasible, NAEP include information on special groups (e.g., information reported by race, ethnicity, socioeconomic status, gender, disability, and limited English proficiency).

As part of most NAEP assessments, three types of questionnaires are used to collect information: student, teacher, and school. An overview of the questionnaires is presented below, and additional information about the content of the questionnaires is presented in Part C.

Student Questionnaires

Each NAEP student assessment booklet includes non-cognitive items, also known as the student questionnaire. The questionnaires appear in separately timed blocks of items in the assessment forms. The items collect information on students' demographic characteristics, classroom experiences, and educational support. Students' responses provide data that give context to NAEP results and/or allow researchers to track factors associated with academic achievement. Students complete the questionnaires voluntarily, and their responses are kept confidential (see Section A.10 for more information). Student names are never reported with their responses or with the other information collected by NAEP.

Each student questionnaire includes three types of items:

- General student information: Student responses to these items are used to collect information about factors such as race or ethnicity and parents' education level. Answers on the questionnaires also provide information about factors associated with academic performance, including homework habits, the language spoken in the home, and the number of books in the home.
- Other contextual/policy information: These items focus on students' educational settings and experiences, and collect information about students' attendance (i.e., days absent), family discourse (i.e., talking about school at home), reading load (i.e., pages read per day), and exposure to English in the home. There are also items that ask about students' effort on the assessment, and the difficulty of the assessment. Answers on the questionnaires provide information on how aspects of education and educational resources are distributed among different groups.
- Subject-specific information: In most NAEP administrations, these items cover three categories of information: (1) time spent studying the subject; (2) instructional experiences in the subject; and (3) student factors (e.g., effort, confidence) related to the subject and the assessment.

Teacher Questionnaires

To provide supplemental information about the instructional experiences reported by students, teachers are asked to complete a questionnaire about their instructional practices, classroom organization, teaching background and training, and the subject in which students are being assessed. Teacher responses are then matched to student data. While completion of the questionnaire is voluntary, NAEP encourages teachers' participation since their responses improve the accuracy and completeness of the NAEP assessment.

Teacher questionnaires are typically only given to teachers at grades 4 and 8; NAEP typically does not collect teacher information for grade 12. By grade 12, there is such variation in student course taking experiences that students cannot be matched to individual teachers for each tested subject. For example, a student may not be taking a mathematics class in grade 12, so he or she cannot be matched to a teacher. Conversely, a student could be taking two reading classes at grade 12 and have multiple teachers related to reading. Only an economics teacher questionnaire has been developed and administrated at grade 12. However, this data was not released (with either the 2006 or the 2012 results) due to a student-teacher match rate below statistical standards².

Teacher questionnaires are organized into different parts. The first part of the teacher questionnaire covers background and general training, and includes items concerning years of teaching experience, certifications, degrees, major and minor fields of study, coursework in education, coursework in specific subject areas, the amount of in-service training, the extent of control over instructional issues, and the availability of resources for the classroom. Subsequent parts of the teacher questionnaire tend to cover training in the subject area,

² The grade 12 economics teacher match rate was 56% in 2012. For comparison, the 2015 teacher match rates for grades 4 and 8 were approximately 94% and 86%, respectively.

classroom instructional information, and teacher exposure to issues related to the subject and the teaching of the subject. They also ask about pre- and in-service training, the ability level of the students in the class, the length of homework assignments, use of particular resources, and how students are assigned to particular classes.

School Questionnaires

The school questionnaire provides supplemental information about school factors that may influence students' achievement. It is given to the principal or another official of each school that participate in the NAEP assessment. While schools' completion of the questionnaire is voluntary, NAEP encourages schools' participation since it makes the NAEP assessment more accurate and complete.

The school questionnaire is organized into different parts. The first part tends to cover characteristics of the school, including the length of the school day and year, school enrollment, absenteeism, dropout rates, and the size and composition of the teaching staff. Subsequent parts of the school questionnaire tend to cover tracking policies, curricula, testing practices, special priorities, and schoolwide programs and problems. The questionnaire also collects information about the availability of resources, policies for parental involvement, special services, and community services.

The supplemental charter school questionnaire designed to collect information on charter school policies and characteristics is provided to administrators of charter schools who are sampled to participate in NAEP. The supplement covers organization and school governance, parental involvement, and curriculum and offerings.

Development of Survey Items

The Background Information Framework and the Governing Board's Policy on the Collection and Reporting of Background Data (located at <https://www.nagb.org/content/nagb/assets/documents/policies/collection-report-backg-data.pdf>), guide the collection and reporting of non-cognitive assessment information. In addition, subject-area frameworks provide guidance on subject-specific non-cognitive assessment questions to be included in the questionnaires. The development process is very similar to the cognitive items, including review of the existing item pool, development of more items than are intended for use, review by experts (including the standing committee), and cognitive interviews with students, teachers, and schools. When developing the questionnaires, NAEP uses a pretesting process so that the final questions are minimally intrusive or sensitive, are grounded in educational research, and the answers can provide information relevant to the subject being assessed.

In the web-based NAEP Data Explorer³, the results of the questionnaires are sorted into eight broad categories: Major Reporting Groups, Student Factors, Factors Beyond School, Instructional Content and Practice, Teacher Factors, School Factors, Community Factors, and Government Factors.

To minimize burden on the respondents and maximize the constructs addressed via the questionnaires, NAEP may spiral items across respondents and/or rotate some non-required items across assessment administrations. The questionnaires are included in Appendix F. This appendix represents a "library" of NAEP items for each subject and respondent. Not all of the items presented would be given to an individual respondent or in a specific administration. In addition, some of the items included in the appendix are being pilot tested in 2016. The data from the pilot will be used to determine the viability of these new items. The final versions of the 2017, 2018, and 2019 questionnaires will be each submitted to OMB as a change request prior to the assessments; these submissions will include a spiral map (if appropriate).

³ See Section A.2 for more information about how NAEP results are reported.

A.1.c.4. Inclusion in NAEP

It is important for NAEP to assess as many students selected to participate as possible. Assessing representative samples of students, including students with disabilities (SD) and English language learners (ELL), helps to ensure that NAEP results accurately reflect the educational performance of all students in the target population, and can continue to serve as a meaningful measure of U.S. students' academic achievement over time.

The National Assessment Governing Board, which sets policy for NAEP, has been exploring ways to ensure that NAEP continues to appropriately include as many students as possible and to do so in a consistent manner for all jurisdictions assessed and reported. In March 2010, the Governing Board adopted a policy, *NAEP Testing and Reporting on Students with Disabilities and English Language Learners* (located at www.nagb.org/content/nagb/assets/documents/policies/naep_testandreport_studentswithdisabilities.pdf). This policy was the culmination of work with experts in testing and curriculum, and those who work with exceptional children and students learning to speak English. The policy aims to:

- maximize participation of sampled students in NAEP,
- reduce variation in exclusion rates for SD and ELL students across states and districts,
- develop uniform national rules for including students in NAEP, and
- ensure that NAEP is fully representative of SD and ELL students.

The policy defines specific inclusion goals for NAEP samples. At the national, state, and district levels, the goal is to include 95 percent of all students selected for the NAEP samples, and 85 percent of those in the NAEP sample who are identified as SD or ELL.

Students are selected to participate in NAEP based on a sampling procedure⁴ designed to yield a sample of students that is representative of students in all schools nationwide and in public schools within each state. First, schools are selected, and then students are sampled from within those schools without regard to disability or English language proficiency. Once students are selected, those previously identified as SD or ELL may be offered accommodations or excluded.

Accommodations in the testing environment or administration procedures are provided for SD and ELL students. Some examples of accommodations permitted by NAEP are extra time, testing in small-group or one-on-one sessions, reading aloud to a student, and scribing a student's responses. Some examples of testing accommodations not allowed are giving the reading assessment in a language other than English, or reading the passages in the reading assessment aloud to the student.

States and jurisdictions vary in their proportions of special-needs students and in their policies on inclusion and the use of accommodations. Despite the increasing identification of SD and ELL students in some states, in particular of ELL students at grade 4, NAEP inclusion rates have generally remained steady or increased since 2003. This reflects efforts on the part of states and jurisdictions to include all students who can meaningfully participate in the NAEP assessments. The new NAEP inclusion policy is an effort to ensure that this trend continues.

A.1.c.5. Transition to Digitally Based Assessments (DBAs)

Virtually all of our nation's schools are equipped with computers, and an increasing number of schools are making digital tools an integral component of the learning environment, reflecting that the knowledge and skills needed for future post-secondary success involve the use of new technologies. NAEP is evolving to address the changing educational landscape through its transition to DBAs. The goal is to be paperless by the end of the decade.

⁴ See Section B.1.a for more information on the NAEP sampling procedures.

NAEP DBAs are using current technology, and as technology evolves, so will the nature of delivery of the assessments. NAEP currently administers the DBAs on tablets, which NAEP field staff bring into the schools⁵. Other administration models may be considered in the future, including the use of school equipment or a combination of approaches.

DBAs allow NAEP to:

- more accurately reflect what is happening in today’s classrooms;
- improve measurement of knowledge and skills; and
- collect new types of data that provide depth in our understanding of what students know and can do, including how they engage with new technologies to approach problem solving.

Approach to the DBA Transition

Given NAEP’s decades of valuable performance information, maintaining trend lines into the future is a high priority. As such, NAEP is using a multistep process to move from paper to digital technology in careful stages that are designed to protect trend reporting. The process involves two stages of piloting before administering an operational NAEP DBA:

- Stage 1 is to adapt the paper-based items for tablet delivery. Comparing results from paper and digitally based versions of the same assessment content administered in the same year allows NAEP to establish a link between administration modes and help its audiences interpret performance trends across the transition from paper to digital delivery.
 - Stage 1 pilots were administered in 2015 for the mathematics, reading, and science assessments.
 - We are studying the mode effect in 2015 and again after 2015 to provide additional information and assurance that NAEP’s trend lines remain meaningful indicators of changes in student performance over time. Paper-based versions of the mathematics and reading assessments will be administered again in 2017 to a portion of the student sample within each state; while the remainder will take the digital version. Inclusion of the paper-based component is designed to support a bridge study that both measures and potentially adjusts the metric in which results are reported for differences due to the change in mode. Details of the bridge study are presented in Section A.1.d.
- Stage 2 is to develop new assessment items and innovative item types and tasks that make use of digital technologies. This new DBA content is gradually introduced into the assessment after first studying the effects of including these new items and item types. In the stage 2 pilots, new items and item types are piloted alongside previously administered items so that the performance of the new items relative to the existing assessment content—and the existing trend line—can be evaluated.
 - The first stage 2 pilots will be given in 2016 in mathematics and reading.

Both stages of piloting are important for ensuring that NAEP’s trend lines can be maintained. For each NAEP subject and grade, the first operational DBA will be composed of the items from the stage 1 pilots and the relatively modest amount of new content from the stage 2 pilots. Over time, more digital content and new item types and tasks will be developed and gradually incorporated into the assessments. Proceeding in this manner helps to ensure that NAEP can continue to meaningfully and reliably report on changes in student performance over time.

Leveraging New Technologies

NAEP’s DBAs will use new testing methods and item types that reflect the growing use of technology in education. Examples of such new item types include:

- Multimedia elements, such as videos and audio clips: The NAEP computer-based writing assessment, administered in 2011 at grades 8 and 12, made use of multimedia. These elements will be incorporated

⁵ See Section B.2 regarding procedures for data collection.

into other NAEP DBAs as well. The 2011 writing tasks were presented to students on computers in a variety of ways, including text, audio, photographs, video, and animation. Examples of these tasks are available at http://www.nationsreportcard.gov/writing_2011/sample_quest.aspx.

- Interactive items and tools: Some questions may allow the use of embedded technological features to form a response. For example, students may use “drag and drop” functionality to place labels on a graphic, or may tap an area or zone on the screen to make a selection. Other questions may involve the use of digital tools. In the mathematics DBA, an online calculator is available for students to use when responding to some items. An equation editor is also provided for the entry of mathematical expressions and equations, and we are exploring the development of other tools, such as digital rulers and protractors, that can be used to gauge students’ mathematics skills. Students are shown how to use these interactive features and tools in the brief tutorials that are included at the beginning of each NAEP DBA.
- Immersive scenario-based tasks: Scenario-based tasks use multimedia features and tools to engage students in rich, authentic problem solving contexts. NAEP’s first scenario-based tasks were administered in 2009, when students at grades 4, 8, and 12 were assessed with interactive computer tasks in science. The science tasks asked students to solve scientific problems and perform experiments, often by simulation. They provide students more opportunities than a paper based assessment (PBA) to demonstrate skills involved in doing science without many of the logistical constraints associated with a natural or laboratory setting. The science tasks administered in 2009 can be explored at http://www.nationsreportcard.gov/science_2009/ict_summary.aspx. NAEP also administered scenario-based tasks in the 2014 technology and engineering literacy (TEL) assessment, where students were challenged to work through computer simulations of real-world situations they might encounter in their everyday lives. A sample TEL task can be viewed at http://nces.ed.gov/nationsreportcard/tel/wells_item.aspx. NAEP is exploring the use of scenario-based tasks to measure knowledge and skills in other subject areas, such as mathematics and reading.

In addition to new types, the transition to DBAs makes it possible for NAEP to employ an adaptive testing design, in which assessment content is targeted to a student’s ability based on performance during the test administration. Thus, students see items that are tailored to their ability levels, and they may be more likely to be able to engage in the assessment and demonstrate what they know and can do. The goal of implementing adaptive testing is to achieve better measurement of student knowledge and skills across the wide range of student performance levels in which NAEP reports. NAEP is considering using adaptive testing initially in the mathematics and science DBAs and possibly in other NAEP assessments in the future.

The type of adaptive testing being considered for NAEP is a multi-stage test (MST) design. There would be two stages. Students would take two sections of items, just as in NAEP PBAs. Based on performance on the first section of items, students would receive a second section of items that is targeted to their ability level. For example, students who do not perform well on the first section of items would receive a second section composed of somewhat easier items. The implementation of this two-stage MST design for NAEP mathematics and science has been informed by previous research on the benefits, applicability, and feasibility of adaptive testing for NAEP. In particular, in 2011 NAEP conducted the mathematics computer-based study, which evaluated the use of a two-stage MST design for the grade 8 mathematics assessment⁶. In addition, the 2015 stage 1 pilots in mathematics and science also incorporated an MST design.

These new item types and testing technologies may allow NAEP to capture information about students’ problem solving processes and the strategies they use to answer items. For example, while PBAs would only yield the final responses in the test booklet, DBAs capture information about student use of the tools,

⁶ The study design and results are summarized in Oranje, A., Mazzeo, J., Xu, X., & Kulick, E. (2014). A multistage testing approach to group-score assessments. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 371-389). Boca Raton, FL: CRC Press.

whether students change their answer, etc. As such, NAEP will potentially uncover more information about which skills successful students use and where the skills of less successful students break down.

Development of Digitally Based Assessments (DBAs)

NAEP's item and system development processes include several types of activities that help to ensure that our DBAs measure the subject-area knowledge and skills outlined in the NAEP frameworks and not students' ability to use the tablet or the particular software and digital tools included in the DBAs.

During item development, new digitally based item types and tasks are studied and pretested with diverse groups of students. The purpose of these pretesting activities is to determine whether construct-irrelevant features, such as confusing wording, unfamiliar interactivity or contexts, or other factors, prevent students from demonstrating the targeted knowledge, skills, and abilities. Such activities help identify usability, design, and validity issues so that items and tasks may be further revised and refined prior to administration.

Development of the assessment delivery system, including the interface that students interact with when taking NAEP DBAs, is informed by best practices in user experience design. Decisions about the availability, appearance, and functionality of system features and tools are also made based on the results of usability testing with students.

To help ensure that students know how to use the assessment system and tools, each administration of a NAEP DBA begins with a brief interactive tutorial that teaches students how to use the system features to take the assessment. Students actively engage with the tutorial, as they are asked to use specific tools and features. Help screens are also built into the system, and students can access them at any time while taking the assessment.

Videos of the tutorials used in recent DBAs are available on the NAEP website at <https://nces.ed.gov/nationsreportcard/dba/>

Accommodations and universal design features are offered with DBAs

New technologies are improving NAEP's ability to offer accommodations to increase participation and provide universal access to students of all learning backgrounds, including students with disabilities and English language learners. In a digital environment, what used to be an accommodation for PBAs becomes a seamless part of universal design, available to all students. This means that things like adjusting font size, having test items read aloud in English (text-to-speech), changing the appearance of the testing interface to have a higher and a lower contrast, using a highlighter tool, and marking answer choices to eliminate them before selecting a final choice can be accomplished by all students during the test administration.

In addition to these universal design features, NAEP also continues to offer accommodations to students with IEPs and 504 plans requiring that they have them. Some accommodations are available in the testing system (such as additional time or a magnification tool), while others are provided by the test administrator or the school (such as breaks during testing or sign language interpretation of the test). Section B.2 b provides more information on the classification of students and the assignment of accommodations.

A.1.c.6. Assessment Types

NAEP uses three types of assessment activities, which may simultaneously be in the field during any given data collection effort. Each is described in more detail below.

Operational assessments

“Operational” NAEP administrations, unlike pilot administrations, collect data to publicly report on the educational achievement of students as required by Federal law. The NAEP results are reported in the Nation's Report Card (<http://nationsreportcard.gov/>), which is used by policymakers, state and local educators, principals, teachers, and parents to inform educational policy decisions.

Pilot assessments

Pilot testing (also known as field testing) of cognitive and non-cognitive items is carried out in all subject areas. Pilot assessments are conducted in conjunction with operational assessments and use the same procedures as the operational assessments. The purpose of pilot testing is to obtain information regarding clarity, difficulty levels, timing, and feasibility of items and conditions. In addition to ensuring that items measure what is intended, the data collected from pilot tests serve as the basis for selecting the most effective items and data collection procedures for the subsequent operational assessments. Pilot testing is a cost-effective means for revising and selecting items prior to an operational data collection because the items are administered to a small nationally representative sample of students and data are gathered about performance that crosses the spectrum of student achievement. Items that do not work well can be dropped or modified before the operational administration.

Prior to pilot testing, many new items are pre-tested with small groups of sample participants (cleared under the NCES pretesting generic clearance agreement; OMB #1850-0803). All non-cognitive items undergo one-on-one cognitive interviews, which is useful for identifying questionnaire and procedural problems before larger scale pilot testing is undertaken. Select cognitive items also undergo pre-pilot testing, such as item tryouts or cognitive interviews, in order to test out new item types or formats, or challenging content. In addition, usability testing is conducted on new technologies and technology based platforms and instruments.

Special studies

Special studies are an opportunity for NAEP to investigate particular aspects of the assessment without impacting the reporting of NAEP results. Previous special studies have focused on linking NAEP to other assessments or linking across NAEP same subject frameworks, investigating the expansion of the item pool, evaluating specific accommodations, investigating administration modes (such as DBA alternatives), and providing targeted data on specific student populations.

In addition to the overarching goal of NAEP to provide data about student achievement at the national, state, and district levels, NAEP also provides specially targeted data on an as-needed basis. At times, this may only mean that a special analysis of the existing data is necessary. At other times, this may include the addition of a short add-on questionnaire targeted at specified groups. For example, in the past, additional student, teacher, and school questionnaires were developed and administered as part of the National Indian Education Study (NIES) that NCES conducted on behalf of the Office of Indian Education. Through such targeted questionnaires, important information about the achievement of a specific group is gathered at minimal additional burden. These types of special studies are intentionally kept to a minimum and are designed to avoid jeopardizing the main purpose of the program.

A.1.d. Overview of 2017-2019 NAEP Assessments

The Governing Board determines NAEP policy and the assessment schedule⁷, and future Governing Board decisions may result in changes to the plans represented here. Any changes will be presented in subsequent clearance packages or revisions to the current package.

The 2017 data collection will consist of the following:

- Operational national, state (including Puerto Rico⁸), and TUDA DBAs in reading and mathematics at grades 4 and 8;
- Operational national DBAs in writing at grades 4 and 8;

⁷ The Governing Board assessment schedule can be found at <http://www.nagb.org/naep/assessment-schedule.htm>.

⁸ Puerto Rico is administered a Spanish-language version of the mathematics assessment. Puerto Rico does not participate in the NAEP reading assessment because the assessment measures a student's ability to read in English.

- Pilot DBAs for 2019 reading and mathematics at grades 4 and 8;
- Pilot⁹ DBAs for 2018 U.S. history, civics, and geography at grade 8;
- PBAs, state (including Puerto Rico), and TUDA bridge studies in reading and mathematics at grades 4 and 8;
- Computer access and familiarity study at grades 4 and 8;
- Multi-stage testing study in mathematics at grades 4 and 8;
- Knowledge and skills appropriate study in mathematics at grades 4 and 8; and
- Laptop bridge study in writing at grade 8 (administered after the regular NAEP assessment window).

The 2018 data collection will consist of the following:

- Operational national DBAs in U.S. history, civics, and geography assessments at grade 8;
- Operational national DBAs¹⁰ in TEL at grade 8;
- Pilot DBAs for 2019 reading and mathematics at grade 12;
- Pilot DBAs for 2019 science at grades 4, 8, and 12; and
- PBA bridge studies in U.S. history, civics, and geography at grade 8.

The 2019 data collection will consist of the following:

- Operational national, state (including Puerto Rico), and TUDA DBAs in reading and mathematics at grades 4 and 8;
- Operational national DBAs in reading and mathematics at grade 12;
- Operational national DBAs in science at grades 4, 8, and 12;
- Pilot DBAs in reading, mathematics, and writing at grades 4, 8, and 12;
- High School Transcript Study; and
- National Indian Education Study.

The planned special studies are conducted in accordance with the assessment development, research, or additional reporting needs of NAEP. With the exception of the High School Transcript Study and the National Indian Education Study, all data collection procedures are the same as those for operational and pilot NAEP assessments (as described in Part B.2). Additional details for the High School Transcript Study and the National Indian Education Study will be provided in 2018 (prior to these studies being conducted in 2019). At that point NCES will (a) publish on Regulations.gov an amendment to this package with all details for these special studies, (b) announce a 30-day public comment period on these details in the Federal Register, and (c) submit the amendment to OMB for review. Additional details on each of the special studies are provided below.

High School Transcript Study (HSTS)

Through the NAEP High School Transcript Study (HSTS), the National Center for Education Statistics (NCES), periodically surveys the curricula being followed in our nation's high schools and the course-taking patterns of high school students through a collection of transcripts. Conducted in conjunction with NAEP, HSTS also offers information on the relationship of student course-taking patterns to achievement at grade 12 as measured by NAEP. With the most recently reported 2009 study, HSTS provides over a decade of valuable findings to the education community.

The 2009 transcript study was conducted from late spring through the January 2010 after the administration of NAEP. Transcripts were collected for twelfth-grade students who graduated high school by the end of the collection period. Most students also participated in the NAEP assessments earlier that same year.

NAEP-related transcript studies were also conducted in previous years. In addition to the 2009 transcript study, the study was also conducted in 1987, 1990, 1994, 1998, 2000, and [2005](#). The 2019 HSTS study, will

⁹ The 2017 pilot DBAs in U.S. history, civics, and geography include both Stage 1 and Stage 2 pilots (see Section A.1.c.5).

¹⁰ While all other DBAs are administered on tablet, the 2018 TEL will be administered on laptop for comparability with the 2014 TEL.

be conducted at approximately 800 schools, and will utilize similar methods as those used in previous years. As noted above, an amendment to this package describing the study details will be submitted for approval prior to conducting the study. Information related to the sampling, design, data collection methods, and analyses, as well as results from previous studies, can be found at <http://nces.ed.gov/nationsreportcard/hsts/>.

National Indian Education Survey (NIES)

The National Indian Education Study (NIES) is designed to describe the condition of education for American Indian and Alaska Native (AI/AN) students in the United States. The study provides educators, policymakers, and the public with information about the academic performance in reading and mathematics of AI/AN fourth- and eighth-graders as well as their exposure to Native American culture and language.

Conducted in conjunction with the NAEP assessments in 2005, 2007, 2009, 2011, and 2015, NIES provides data on a nationally representative sample of American Indian and Alaska Native students in public, private, Department of Defense, and Bureau of Indian Education funded schools. It is an important source of data on American Indian and Alaska Native students, especially for educators, administrators, and policymakers who address the educational needs of these students.

The study is sponsored by the Office of Indian Education (OIE) and conducted by NCES for the U.S. Department of Education. A Technical Review Panel (see Appendix A-4), whose members include American Indian and Alaska Native educators and researchers from across the country, help design the study.

This study was conducted through a survey to explore the educational experiences of the fourth- and eighth-grade American Indian and Alaska Native students based on responses to the NIES student, teacher, and school questionnaires. The survey focused on the integration of native language and culture into school and classroom activities.

The 2019 NIES study will use similar methods as those used in previous years. Approximately 8,000 fourth-grade and 6,500 eighth-grade students will participate in the 2019 NIES study. As noted above, an amendment to this package describing the study details will be submitted for approval prior to conducting the study. Information related to the sampling, design, data collection methods, and analyses, as well as results from previous studies can be found at <http://nces.ed.gov/nationsreportcard/nies/>.

Computer Access and Familiarity Study (CAFS)

As NAEP transitions from PBAs to DBAs, an area of desired research involves the degree to which all children are ready for such a transition. Do all students have the same access and experience with the technologies (computers and tablets) that will be used to collect the data? What is the relationship between access and experience with these technologies and performance on NAEP assessments? The study will analyze a core set of items to measure access to, and familiarity with, in relation to the DBA equipment that has been used by NAEP or might be used for future NAEP assessments. The goal is to build reliable composites that measure technology access and familiarity. The study contains a supplemental survey questionnaire related to computer familiarity and access. This study will be the second iteration of the study conducted in 2015.

The 2017 CAFS sample will be a nationally representative subsample of 150 public schools participating in the reading and mathematics operational assessments at grades 4 and 8. The sample will be stratified on characteristics such as census region, urban/rural, school race/ethnicity composition, and school enrollment size. All NAEP sampled students in the subsample of schools will participate in the CAFS study. Within a school selected for the NAEP reading and mathematics assessments, students will be randomly assigned to either DBA or PBA. The ratio of sample sizes for the two modes within each school will be approximately 4:1, with some variation depending upon the size of the school and the jurisdiction.

The expected yield is approximately 3,000 DBA students per grade/subject and 750 PBA students per grade/subject. Based on the results of the 2015 study, it was determined that a minimum sample size of 750 students were needed for each grade, subject, and mode. This sample size supports sufficient power in detecting an effect size of 0.1-0.16 and 0.2-0.32 for DBA and PBA, respectively, between students with low and high computer familiarity. This means that a sample of 150 schools per grade is needed to provide this sample size of 750 students per subject for PBA. These schools will also contain 3,000 students per subject who will be assessed using DBA. It is highly desirable from an operational perspective to have all NAEP students in a school complete the CAFS questionnaire, rather than a subset, and having the additional DBA sample will provide additional power for certain analyses. For the PBA sample, after the students complete their regular printed NAEP booklet, they will be given a separate booklet of CAFS questions. For the DBA sample, the CAFS questions will be an additional section of the student questionnaire, which is administered on the tablet.

Some analyses will be conducted combining the students in the different NAEP subjects, while other analyses will focus within subject only. Analyses, including factor analyses, IRT scaling, and correlational analyses, will examine the relationship between access and familiarity and performance on NAEP (overall and for certain subgroups), if the relationship varies by subject area or mode of administration, and if reliable composites related to computer access and familiarity can be constructed. The goal of the study is to inform the development and use of computer access and familiarity items in the questionnaires and reports for future NAEP assessment years.

Multi-Stage Testing (MST) Study

As described in Section A.1.c.5, NAEP is considering incorporating MST in NAEP DBAs. Prior to implementing MST on an operational-level, NCES will study the implementation of an MST design on the methodologies and results, similar to the study conducted in 2011. The 2011 study was exploratory and a necessary first step to examine potential gains of MST for the NAEP program before a much more significant investment for operational deployment (in terms of resources, reputation, trend maintenance) could be considered. Gains were defined in terms of (conditional) standard errors, ability to meaningfully describe performance over a wider range of proficiency levels, and student engagement. A subset from the existing and pilot item pool containing predominantly multiple-choice, paper-based items, were transformed for computer-based assessment administered to an approximately national sample.

The current study is entirely geared towards preparing for operational deployment using a subset of items from the 2017 operational pool, the operational delivery system on tablets, and a nationally representative sample. The advisability to study an operational design before deploying operationally rests on the fact that, at the very core, the NAEP program is charged with maintaining trends. Therefore, any significant design changes require careful study and, in many cases, carefully designed bridge studies, in order not to interfere with the ability to maintain a robust trend. Given that much of the previous research on MST design and implementation has been conducted on individual assessments and the psychometric and statistical parameters are very different for individual assessments than group-score assessments (such as NAEP), it is critical to study this major design change in the NAEP setting.

The 2017 MST study will be conducted at both grades 4 and 8 mathematics in conjunction with the operational assessments. As such, the same sampling, recruitment, and administration procedures as the operational assessments will be used. The only difference between this study and the operational assessment is how items are assigned to blocks and how blocks are assigned to students. In this study, students will first be randomly assigned to a 30-minute routing block, and then routed to a second 30-minute block targeted to ability level: easy, medium, or hard. The second stage (i.e., the target block) has different designs for the two grades. For grade 4, the design includes an adjacent routing component where some students are assigned to the adjacent targeted level rather than their intended level (i.e., some students routed to easy will be assigned a medium block). There will be no overlapping of items across blocks. For grade 8, on the other

hand, blocks will be assembled with overlapping items among routers and between targeted levels. However, there is no adjacent routing component at grade 8 (therefore, all students routed to an easy block will be assigned an easy block). The analysis will evaluate the IRT item parameter estimates obtained from the MST designs in relationship to the item parameter estimates from the 2017 operational DBAs as the baseline. Consistency in parameter estimates between the 2017 MST study and the operational DBAs would be a positive outcome, indicating the MST design can be implemented in NAEP going forward.

The 2017 MST study will be administered to a national sample of 10,000 students at each grade. As with operational assessments, the sample size for this special study is primarily driven by the need for sufficient numbers of student responses at each item to support IRT calibration. For grade 4, the target sample size is approximately 3,000 per item for the first stage routing blocks, and approximately 1,100 to 2,700 per item for the second stage target blocks. For grade 8, the target sample size is approximately 3,300 and 800 per item for the first stage routing blocks and the second stage target blocks, respectively. The variation in sample sizes are functions of different numbers of blocks at each stage, as well as at each targeted level.

Knowledge and Skills Appropriate (KaSA) in Mathematics

NAEP has had difficulty measuring the abilities of lower-performing students in jurisdictions such as Puerto Rico. In an effort to obtain more information on what low-performing students in jurisdictions such as Puerto Rico know and can do, new fourth- and eighth-grade mathematics items were developed to be more knowledge and skills appropriate (KaSA) for such students. Administered in conjunction with the NAEP mathematics assessments in 2011, 2013, and 2015, KaSA allows for scores from Puerto Rico to be placed on the NAEP scale.

While the original KaSA instrument was designed to address a broader need to improve measurement precision on low-performing students, the KaSA special study has only been implemented in Puerto Rico as NAEP has had difficulties historically reporting scale scores for Puerto Rico. As the program moves to multi-stage testing design, KaSA items will be part of the MST instrument. And the selection of students (from all jurisdictions, including Puerto Rico) receiving KaSA items, as well as other targeted items, will be based on their performance on the routing items.

Currently, the KaSA special study serves as a bridge to enable NAEP to report on Puerto Rico similar to other jurisdictions. The 2017 KaSA study will be conducted at both grades 4 and 8 in conjunction with the operational mainland assessments. As such, the same sampling, recruitment, and administration procedures as the operational assessments will be used. For each administration mode (PBA and DBA) in 2017, the study design involves both Puerto Rico sample (3,000) and a nationally representative linking sample (3,000) receiving KaSA blocks in addition to the operational blocks. The sample sizes are primarily driven by the need for sufficient numbers of student responses per item to support IRT item calibration, as well as to support Puerto Rico jurisdiction-level reporting. During analysis, a statistical linking approach in IRT calibration is used to link the Puerto Rico student proficiency onto the operational reporting scale. Using this KaSA special study methodology, NAEP has been able to report scale scores for Puerto Rico since 2011.

Digitally Based Assessment (DBA) Bridge Studies

The term “bridge study” is used to describe a study conducted so that the interpretation of the assessment results remains constant over time. A bridge study involves administering two assessments: one that replicates the assessment given in the previous assessment year using the same questions and administration procedures (a bridge assessment), and one that represents the new design (a modified assessment). Comparing the results from the two assessments, given in the same year to randomly equivalent groups of students, provides an indication of whether there are any significant changes in results caused by the changes in the assessment. A statistical linking procedure can then be employed, if necessary, to adjust the scores so they are on the same metric, allowing trends to be reported. Three DBA bridge studies are planned:

- In 2017, PBA bridge studies are planned in reading and mathematics in addition to the operational DBAs to confirm the findings from the 2015 initial national-level bridge studies;
- In 2018, a PBA to DBA bridge study is planned in U.S. history, civics, and geography; and
- In 2017, a laptop to tablet DBA comparability bridge study is planned in writing at grade 8; it will be conducted after the regular NAEP administration window.

As described in A.1.c.5, NAEP is using a multi-step process designed to protect trend reporting to transition from PBA to DBA. For reading and mathematics at grades 4 and 8, the 2015 PBAs will be re-administered at the state and TUDA level in 2017, along with the operational DBAs.

In 2017, the PBAs will be administered to a representative sample in each jurisdiction, enabling the examination of the relationship between PBA and DBA performance within each jurisdiction. The targeted PBA sample size is 500 students per state and TUDA, as well as 500 private school students for each subject within a grade. The sample sizes are driven by the need for sufficient numbers of student responses per item to support IRT item calibration, as well as to support evaluating mode effect at the state and TUDA level and for the private school population. The PBA will allow us to both measure and potentially adjust for differences due to the change in mode.

Similar PBA bridge studies will be conducted in 2018 for U.S. history, civics, and geography at grade 8. Given that the operational assessments of those three subjects are at the national-level, in 2018, the PBA will be administered to a nationally representative sample for each subject. The total sample size across the three subjects is 24,000. The size of national sample is primarily driven by the need for sufficient numbers of student responses at item level to support IRT calibration.

In addition to the PBA to DBA bridge studies mentioned above, NAEP will also study the transition from laptop-administration to tablet administration in writing. The first operational writing DBA was administered on laptop in 2011 at grades 8 and 12. The grade 8 writing assessment will shift delivery mode from laptop to tablet for the 2017 operational administration (note, grade 12 is not being administered in 2017). The goal of this study is to gather information about potential device effects on grade 8 student writing performance on tasks. Student writing performance on tasks on two devices—tablet vs. laptop—will be compared. This information will support better interpretation of trend results between 2017 and 2011.

A nationally representative sample of 3,000 students will participate in the study. Although the study will be administered during a separate window (from April to May 2017, as opposed to the rest of NAEP being administered from January to March 2017), the recruitment, sampling, and administration procedures described in Sections B.2.a, B.1.a, and B.2.c, respectively, will be used in this study. Six of the writing tasks administered as part of the 2017 operational tablet-based grade 8 assessment will also be administered in this study. The task level summary statistics (e.g., average score on a writing task) will be compared, along with score distributions (ranging from 0 to 5). The comparison information will be used to inform interpretation of the trend results between 2017 and 2011.

While the sample size for most NAEP assessments is primarily driven by the need for sufficient numbers of student responses per item to support IRT item calibration, no item calibration is planned for the laptop-based sample. Therefore, the sample size for this study supports sufficient power (at least 0.8 with significant level of 0.05) in detecting a small effect size of 0.2 for average task score comparisons between the two devices.

A.2. How, by Whom, and for What Purpose the Data Will Be Used

Results will be reported on the 2017 operational assessments in mathematics, reading, and writing; the 2018 operational assessments in TEL, U.S. history, geography, and civics; and the 2019 operational assessments in mathematics, reading, and science. In addition, the DBA bridge studies will be used to inform the operational DBA results. Results will also be reported from the 2019 HSTS and NIES special studies. NAEP will use the results from the pilot tests to inform future assessments and procedures.

The NAEP operational results are reported in the Nation's Report Card, which is used by policymakers, state and local educators, principals, teachers, and parents to help inform educational policy decisions. The NAEP Report Cards provide national results, trends for different student groups, results on scale scores and achievement levels, and sample items. In reports with state or urban district results, there are sections that provide overview information on the performance of these jurisdictions. NAEP does not provide scores for individual students or schools.

Results from each NAEP assessment are provided online in an interactive website (<http://nationsreportcard.gov/>) and in one-page summary reports, called snapshots, for each participating state or urban district. Additional data tools are available online for those interested in:

- analyzing NAEP data and creating tables and graphics (<http://nces.ed.gov/nationsreportcard/naepdata/>);
- comparing state performance by various demographic groups (<http://nces.ed.gov/nationsreportcard/statecomparisons/>);
- seeing NAEP performance results and student demographics for each state (<http://nces.ed.gov/nationsreportcard/states/>);
- browsing results for each of the participating large urban districts (<http://nces.ed.gov/nationsreportcard/districts/>);
- searching, sorting, and providing data for sample NAEP items (<http://nces.ed.gov/nationsreportcard/itmrlsx/>); and
- seeing the knowledge and skills demonstrated by students performing at different scale scores (<http://nces.ed.gov/nationsreportcard/itemmaps/>).

In addition to contributing to the reporting tools mentioned above, data from the questionnaires are used as part of the marginal estimation procedures that produce the student achievement results. Questionnaire data is also used to perform quality control checks on school-reported data and in special reports, such as the *Black–White Achievement Gap* report (<http://nces.ed.gov/nationsreportcard/studies/gaps/>).

Lastly, there are numerous opportunities for secondary data analysis because of NAEP's large scale, the regularity of its administrations, and its stringent quality control processes for data collection and analysis. NAEP data are used by researchers and educators who have diverse interests and varying levels of analytical experience.

A.3. Improved Use of Technology

NAEP has continually moved to administration methods that include greater use of technology, as described below.

Online Teacher and School Questionnaires

The teacher and school questionnaires that accompany the NAEP assessment were traditionally available as paper-based questionnaires. Starting in 2001, NAEP offered teachers and school administrators an option of either completing the questionnaires on paper or online. In an effort to reduce costs and to streamline the data collection, starting in 2014 the NAEP program moved to the practice of having the teacher and school questionnaires available primarily online. To support respondents who have limited internet connections, NAEP field staff have limited number of printed copies of the questionnaires that can be distributed at the school's request.

Electronic Pre-Assessment Activities

Each school participating in NAEP has a designated staff member to serve as its NAEP school coordinator. Pre-assessment and assessment activities include functions such as finalizing student samples, verifying student demographics, reviewing accommodations, and planning logistics for the assessment. NAEP is moving in the direction of paperless administrations. An electronic pre-assessment system (known as MyNAEP) was developed so that school coordinators would provide requested administration information online, including logistical information, updates of student and teacher information, and the completion of inclusion and accommodation information¹¹.

Digitally Based Assessments (DBAs)

As described in Section A.1.c.5, NAEP is transitioning to DBA. The move to DBA will allow NAEP to provide assessments consistent with other large-scale assessments (such as those given by the Partnership for Assessment of Readiness for College and Careers [PARCC] and the Smarter Balanced Assessment Consortium). In addition, the transition to DBA allows NAEP to more accurately reflect what is happening in today's classrooms, improve measurement of knowledge and skills, and collect new types of data that provide depth in our understanding of what students know and can do.

Automated Scoring

NAEP administers a combination of selected-response items and open-ended, or constructed-response items. NAEP currently uses human scorers to score the constructed response items, using detailed scoring rubrics and proven scoring methodologies. With the increased use of technologies, the methodology and reliability of automated scoring (i.e., the scoring of constructed-response items using computer software) has advanced. While NAEP does not currently employ automated scoring methodologies, these may be investigated and ultimately employed during the assessment period of 2017-2019.

One possible study involves using two different automated scoring engines and comparing the scores to those previously given by human scorers. This study would be conducted on items from the 2011 writing assessment, as well as some items from the 2015 DBA pilot. For each constructed response item, approximately two-thirds of responses would be used to develop the automated scoring model (the Training/Evaluation set) and the other third of responses would be used to test and validate the automated scoring model (the Test/Validation set).

The Training/Evaluation set would be used to train, evaluate, and tune each scoring engine so as to produce the best possible scoring models for each constructed response item. The final scoring models would then be applied to the Test/Validation set producing a holistic score for each response.

Automated scoring performance is typically evaluated by comparison with human scoring performance. Evaluation criteria for the scoring models would include measures of inter-rater agreement such as correlation, quadratic-weighted kappa, exact and adjacent agreement, and standardized mean difference.¹²

¹¹ Additional information on the MyNAEP site is included in the Section B.2.

¹² Evaluation criteria will be based on criteria advocated in Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practices*, 31(1), 2-13.

These measures would be computed for pairs of human ratings as well as for pairs of automated and human scores.

In addition to comparing how well each individual scoring engine agrees with human scorers, we would also compute how well the two scoring engines agree with each other, and how well the combination of the two engines (computed by averaging their scores) agrees with the human scorers. Results of these investigations would determine if automated scoring could be utilized for specific NAEP assessments or if additional investigations are required.

A.4. Efforts to Identify Duplication

The proposed assessments, including the questionnaires, do not exist in the same format or combination in the U.S. Department of Education or elsewhere. The non-cognitive data gathered by NAEP comprise the only comprehensive cross-sectional survey performed regularly on a large-scale basis that can be related to extensive achievement data in the United States. No other federally funded studies have been designed to collect data for the purpose of regularly assessing trends in educational progress and comparing these trends across states. None of the major non-federal studies of educational achievement were designed to measure changes in national achievement. In short, no existing data source in the public or private sector duplicates NAEP.

While the survey items in NAEP are unique, the items are not developed in a vacuum. Their development is informed by similar items in other assessments and survey programs. In addition, in future rounds of development, NCES will continue to better align the NAEP survey questions with other surveys (particularly, but not limited to, those from other NCES and federal survey programs).

Historically, NAEP has served as a critical national "audit" function, offering an extremely helpful reference point in the interpretation of score trends on "high-stakes" tests used for school accountability. The main NAEP scales have served this function well even though high-stakes state assessments were not always closely aligned with the corresponding NAEP assessments. Given the significant changes currently underway in the American educational landscape, including the Next Generation Science Standards, the Common Core State Standards, and Partnership for Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced consortia, this "audit" function is even more important.

NAEP has provided the best available information about the academic achievement of the nation's students in relation to consensus assessment frameworks, maintaining long-term trend lines for decades. In addition to reporting at the national level, NAEP has offered achievement comparisons among participating states for more than two decades, and since 2003, all states have participated in the NAEP mathematics and reading assessments at the fourth and eighth grades. More recently, NAEP has also reported achievement for selected large urban school districts. In addition to characterizing the achievement of fourth-, eighth-, and twelfth-grade students in a variety of subject areas, NAEP has also served to document the often substantial disparities in achievement across demographic groups, tracking both achievement and achievement gaps over time. In addition to describing educational achievement, NAEP has furthered deliberation as to the scope and meaning of achievement in mathematics, reading, and other subject areas. NAEP assessments are aligned to ambitious assessment frameworks developed by a thoughtful process to reflect the best thinking of educators and content specialists. These frameworks have served as models for the states and other organizations to follow. Finally, NAEP has also served as a laboratory for innovation, developing and demonstrating new item formats, as well as statistical methods and models now emulated by large-scale assessments worldwide.

NAEP has functioned well as a suite of complex survey modules conducted as assessments of student achievement in fixed testing windows. The complexity of NAEP evolved by necessity to address its legal and policy reporting requirements and the complex sampling of items and students needed to make reliable

and valid inferences at the subgroup, district, state, and national level for stakeholders ranging from policymakers to secondary analysts, and do so without creating an undue burden on students and schools.

A.5. Burden on Small Businesses or Other Small Entities

The school samples for NAEP contain small-, medium-, and large-size schools, including private schools. Schools are included in the sample proportional to their representation in the population, or as necessary to meet reporting goals. It is necessary to include small and private schools so that the students attending such schools are represented in the data collection and in the reports. The trained field staff work closely with all schools to ensure that the pre-assessment activities and the administration can be completed with minimal disruption.

A.6. Consequences of Collecting Information Less Frequently

Under the National Assessment of Educational Progress Authorization Act, Congress has mandated the ongoing collection of NAEP data. Failure to collect the 2017–2019 assessment data on the current schedule would affect the quality and schedule of the NAEP assessments, and would result in assessments that would not fulfill the mandate of the legislation.

A.7. Consistency with 5 CFR 1320.5

No special circumstances are involved. This data collection observes all requirements of 5 CFR 1320.5.

A.8. Consultations Outside the Agency

The NAEP assessments are conducted by an alliance of organizations under contract with the U.S. Department of Education¹³. The Alliance includes the following:

- Business Intelligence, Inc. is responsible for managing the integration of multiple NAEP project schedules and providing data on timeliness, deliverables, and cost performance.
- Educational Testing Service (ETS) is responsible for coordinating Alliance contractor activities, developing the assessment instruments, analyzing the data, and preparing the reports.
- Fulcrum is responsible for NAEP web operations and maintenance and the development of NAEP DBA delivery systems.
- Pearson is responsible for printing and distributing the assessment materials, and for scanning and scoring students' responses.
- Westat is responsible for selecting the school and student samples, and managing field operations.

In addition to the NAEP Alliance, other organizations support the NAEP program, all of which are under contract with the U.S. Department of Education. The current list of organizations¹⁴ include:

- American Institute for Research (AIR) is responsible for providing technical support, conducting studies on state-level NAEP assessments, and running the NAEP Validity Studies Panel.
- Council of Chief State School Officers (CCSSO) is responsible for providing ongoing information about state policies and assessments.
- CRP, Inc. is responsible for providing logistical and programmatic support.
- Hager Sharp is responsible for supporting the planning, development, and dissemination of NAEP publications and outreach activities.
- Human Resources Research Organization (HumRRO) is responsible for performing formative evaluation of the NAEP Alliance activities.
- Optimal Solutions Group is responsible for providing technical support.
- Tribal Tech is responsible for providing support for the National Indian Education Study.

¹³ The current contract expires on March 6, 2018. A new contract will be awarded prior to that date.

¹⁴ The current contracts expire at varying times. As such, the specific contracting organizations may change during the course of the time period covered under this submittal.

In addition to the contractors responsible for the development and administration of the NAEP assessments, the program involves many consultants and is also reviewed by specialists serving on various technical review panels. These consultants and special reviewers bring expertise concerning students of different ages, ethnic backgrounds, geographic regions, learning abilities, and socioeconomic levels; the specific subject areas being assessed; the analysis methodologies employed; and large-scale assessment design and practices. Contractor staff and consultants have reviewed all items for bias and sensitivity issues, grade appropriateness, and appropriateness of content across states.

In particular, subject area standing committees play a central role in the development of NAEP assessment instruments and have been essential in creating assessment content that is appropriate for the targeted populations, and that meets the expectations outlined in the Governing Board frameworks. One of the most important functions of the committees is to contribute to the validation of the assessments. Through detailed reviews of items, scoring guides, tasks, constructed-response item training sets for scorers, and other materials, the committees help establish that the assessments are accurate, accessible, fair, relevant, and grade-level appropriate, and that each item measures the knowledge and skills it was designed to measure. When appropriate, members of subject area standing committees will also review the questionnaires with regards to appropriateness with existing curricular and instructional practices.

Appendix A lists the current members of the following NAEP advisory committees:

- NAEP Design and Analysis Committee
- NAEP Validity Studies Panel
- NAEP Quality Assurance Technical Panel
- NAEP National Indian Education Study Technical Review Panel
- NAEP Civics Standing Committee
- NAEP Economics Standing Committee
- NAEP Geography Standing Committee
- NAEP Mathematics Standing Committee
- NAEP Reading Standing Committee
- NAEP Science Standing Committee
- NAEP Survey Questionnaires Standing Committee
- NAEP Technology and Engineering Literacy Standing Committee
- NAEP U.S. History Standing Committee
- NAEP Writing Standing Committee
- NAEP Principals' Panel Standing Committee
- NAEP Mathematics Translation Review Committee
- NAEP Science Translation Review Committee

As has been the practice for the past few years, OMB representatives will be invited to attend the technical review panel meetings that are most informative for OMB purposes.

In addition to the contractors and the external committees, NCES works with the NAEP State Coordinators, who serve as the liaison between each state education agency and NAEP, coordinating NAEP activities in his or her state. NAEP State Coordinators work directly with the schools selected for NAEP.

A.9. Payments or Gifts to Respondents

In general, there will be no gifts or payments to respondents, although students do get to keep the NAEP pencils or earbuds used in the PBAs and DBAs, respectively. On occasion, NAEP will leave educational materials at schools for their use (e.g., science kits from the science hands-on assessments). Schools participating in the High School Transcript Study are paid the established fee for providing student transcripts. Given that the study pays schools the prevailing rate to perform a standard service, estimates of school-level burden for that function are not included in this volume. Some schools also offer recognition parties with pizza or other perks for students who participate; however, these are not reimbursed by NCES

or the NAEP contractors. If any incentives are proposed as part of a future special study, they would be justified as part of that future clearance package. As appropriate, the amounts would be consistent with amounts approved in other studies with similar conditions.

A.10. Assurance of Confidentiality

NAEP has policies and procedures that ensure privacy, security, and confidentiality in compliance with the legislation (Confidential Information Protection provisions of Title V, Subtitle A, Public Law 107-347 and the National Assessment of Educational Progress Authorization Act). Specifically, for the NAEP project, this ensures that privacy, security, and confidentiality policies and procedures are in compliance with the Privacy Act of 1974 and its amendments, NCES Confidentiality Procedures, and the Department of Education ADP Security manual. The National Assessment of Educational Progress Authorization Act requires the confidentiality of personally identifiable information [20 U.S.C. §9622 (c) (3)]:

(A) IN GENERAL.-- The Commissioner for Education Statistics shall ensure that all personally identifiable information about students, their academic achievement, and their families, and that information with respect to individual schools, remains confidential, in accordance with section 552a of title 5, United States Code.

(B) PROHIBITION.-- The Assessment Board, the Commissioner for Education Statistics, and any contractor or subcontractor shall not maintain any system of records containing a student's name, birth information, Social Security number, or parents' name or names, or any other personally identifiable information.

Each contractor develops a Data Security Plan and NCES ensures that all current contractor policies and procedures are in compliance with all NAEP security and confidentiality requirements. In addition, all NAEP contractor staff with access to confidential NAEP information are required to sign an affidavit of nondisclosure that affirms, under severe penalty for unlawful action, that they will protect NAEP information from non-authorized access or disclosure. The affidavits are in keeping with the NCES Standard for Maintaining Confidentiality (Standard 4-2). All contractors must also comply with directive OM: 5-101, which requires that all staff with access to data protected by the Privacy Act and/or access to U.S. Department of Education systems and who will work on the contract for 30 days or more go through the security screening procedures. In addition, the Sampling and Data Collection (SDC) contractor has obtained from the Department of Education's Chief Information Security Officer (CISO) a Security Authorization to Operate (ATO) at the FISMA Moderate level and adheres to and continuously monitors the security controls in said Authorization. Security controls include secure data processing centers and sites; properly vetted and cleared staff; and data sharing agreements.

An important privacy and confidentiality issue is the protection of the identity of assessed students, their teachers, and their schools. To assure this protection, NAEP has established security procedures, described below, that closely control access to potentially identifying information.

All assessment and questionnaire data are encrypted at all times. This means that NAEP applications that handle assessment and questionnaire data:

- enforce effective authentication password management policies, making it difficult to hack into the data;
- limit authorization to individuals who truly need access to the data, only granting the minimum access to individuals as they need (i.e., least privilege user access);
- keep data encrypted, both in storage and in transport, utilizing volume encryption and transport layer security protocols;
- utilize SSL certificates and HTTPS protocols for web based applications;
- limit access to data via software and firewall configurations as well as not using well known ports for data connections; and

- restrict access to the portable networks utilized to administer an assessment to only assessment devices.

Students' names are submitted to the SDC contractor for selecting the student sample. This list also includes the month/year of birth, race/ethnicity, gender, and status codes for students with disabilities, English language learners, and participation in the National School Lunch Program.

After the student sample is selected, the data for selected students are submitted to the Materials Preparation, Distribution, Processing and Scoring (MDPS) contractor, who includes the data in the packaging and distribution system for the production of student-specific materials (such as labels to attach to the student booklets or log-in ID cards), which are then forwarded to field staff and used to manage and facilitate the assessment. These data are also uploaded to the MyNAEP Prepare for Assessments online system for review by schools and added to the MyNAEP School Control System (SCS) used by field staff to print materials used by the schools. Student information is deleted from the packaging and distribution system after the assessment begins. Student information is deleted from the MyNAEP system typically two weeks after all quality control activities for the assessment are complete.

All paper-based student-specific materials linking Personally Identifiable Information (PII) to assessment materials are destroyed at the schools upon completion of the assessment. The field staff remove names from forms and place the student names in the school storage envelope. The school storage envelope contains all of the forms and materials with student names and is kept at the school until the end of the school year and then destroyed by school personnel¹⁵.

In addition to student information, teacher and principal names are collected and recorded in the MyNAEP Prepare for Assessment online system, which is used to keep track of the distribution and collection of NAEP teacher and school questionnaires. A paper copy of the questionnaire report is printed for use during the assessment, and this paper copy is left in the school storage envelope, which is destroyed at the end of the school year. The teacher and principal names are deleted from the MyNAEP system at the same time the student information is deleted.

For the DBAs, NAEP data are stored on systems in a locked-down environment at a secure hosting facility with strict measures in place to prevent unauthorized online access. The student names are not included on the assessment tablets or stored by the same contractor or on the same database as the student responses. Shortly before, during, and after assessments, assessment data are transmitted through secure, encrypted channels (SSL, SSH) between NAEP systems, the NAEP assessment servers, and the assessment administration devices. Data on those devices are also encrypted—these data can be read only by the assessment software—and the devices are secured against unauthorized use.

Furthermore, to ensure the confidentiality of respondents, NAEP staff will use the following precautions:

- Assessment and questionnaire data files will not identify individual respondents.
- No personally identifiable information, either by schools or respondents, will be gathered or released by third parties. No permanent files of names or other direct identifiers of respondents will be maintained.
- Student participation is voluntary.
- NAEP data are perturbed. Data perturbation is a statistical data editing technique implemented to ensure privacy for student and school respondents to NAEP's assessment questionnaires for assessments in which data are reported or attainable via restricted-use licensing arrangements with NCES. The process is coordinated in strict confidence with the IES Disclosure Review Board (DRB), with details of the process shared only with the DRB and a minimal number of contractor staff.

After the components of NAEP are completed in a school, neither student- nor teacher-reported data are retrievable by personal identifiers. We emphasize that confidentiality is assured for individual schools and

¹⁵ In early May, schools receive an email from the MyNAEP system reminding them to securely destroy the contents of the NAEP storage envelope and confirm that they have done so. The confirmation is recorded in the system and tracked.

for individual students, teachers, and principals. The following text appears on all student assessments and teacher and school questionnaires:

The information you provide will be used for statistical purposes only. In accordance with the Confidential Information Protection provisions of Title V, Subtitle A, Public Law 107-347 and other applicable Federal laws, your responses will be kept confidential and will not be disclosed in identifiable form to anyone other than employees or agents. By law, every NCES employee as well as every agent, such as contractors and NAEP coordinators, has taken an oath and is subject to a jail term of up to 5 years, a fine of up to \$250,000, or both if he or she willfully discloses ANY identifiable information about you.

More specific information about how NAEP handles PII is provided in the table below:

PII is created in the following ways	1. Public and non-public school samples are released by the SDC contractor to NAEP State Coordinators (public schools only), NAEP TUDA Coordinators (public schools only), and SDC Gaining Cooperation Field Staff (non-public schools only) using the secure MyNAEP for Schools web site.
	2. Schools are recruited by SDC field staff for participation in NAEP.
	3. Participating schools need to submit a current roster of students for the sampled grade for student sampling.
	4. Rosters of students can be created by NAEP State Coordinators, NAEP TUDA Coordinators, or NAEP School Coordinators
	a. Rosters are submitted through the secure MyNAEP for Schools web site
	b. Rosters must be in Excel
	5. PII is contained in the roster files: student names, month/year of birth, race/ethnicity, gender, and status codes for students with disabilities, English language learners, and participation in the National School Lunch Program.
PII is moved in the following ways	6. PII is stored in the SDC contractor’s secure data environments.
	1. Student names (PII) are moved to the Materials Preparation, Distribution, Processing and Scoring (MDPS) contractor via a secure FTP site. These names are used to print Student Login Cards
	2. Student Login Cards are only created for students taking DBAs so the student names for the PBA students are not moved
	3. Student PII data is available to the NAEP School Coordinators and the SDC contractor’s Field Staff through the secure MyNAEP for Schools web site.
	a. NAEP School Coordinators can view and update PII for their own schools
	b. NAEP School Coordinators can print materials containing PII for their own schools
	c. NAEP School Coordinators store materials containing PII for their own schools in the “NAEP Secure Storage Envelope”
	d. SDC contractor Field Staff can update PII for schools within their assignment
e. SDC contractor Field Staff can print materials containing PII for schools within their assignment	
f. SDC contractor Field Staff store materials containing PII for schools within their assignment in their “NAEP School Folders”	
PII is destroyed in the following ways	1. MDPS contractor destroys the PII after printing the Student Login Cards
	2. School Coordinators destroy the materials containing PII on or before the end of the school year
	3. SDC contractor Field Staff destroy the materials containing PII after the school assessment has been completed. SDC contractor Field Staff return their NAEP School Folders to Westat Home Office for secure storage, and eventual secure destruction
	4. SDC contractor destroys student names after all weighting quality control checks have been completed, thereby making it impossible to link the responses to any directly identifiable PII. This activity is completed in August (approximately 175 days following the end of the administration).

In addition, parents are notified of the assessment. Appendix D-17 includes a sample parental notification letter regarding NAEP. The letter is adapted for each grade/subject combination and the school principal may edit it. However, the information regarding confidentiality and the appropriate law reference will remain unchanged.

For the HSTS component of NAEP, student transcripts are collected from schools for sampled students, and school staff members complete a School Information Form that provides general information about class periods, credits, graduation requirements, and other aspects of school policy. The HSTS study currently collects transcripts in paper form, and plans to collect electronic transcripts in the future. To maintain the privacy of student and school identities, students' names are removed from the transcripts and questionnaires at the school and given a unique identification number, which is used to match the transcript records to the NAEP questionnaire and performance information, on an individual basis. NCES ensures that the data collected from schools and students are used for statistical purposes only.

A.11. Sensitive Questions

NAEP emphasizes voluntary respondent participation and assures confidentiality of individual responses. Insensitive or offensive items are prohibited by the National Assessment of Educational Progress Authorization Act, and the Governing Board reviews all items for bias and sensitivity. The nature of the questions are guided by the reporting requirements in the legislation, the Governing Board's Policy on the Collection and Reporting of Background Data, and the expertise and guidance of the NAEP Survey Questionnaire Standing Committee (see Appendix A-11). Additional information on the constructs included in the questionnaires is provided in Part C. Throughout the item development process, NCES staff works with consultants, contractors, and internal reviewers to identify and eliminate potential bias in the items.

The NAEP student questionnaires include items that require students to provide responses on factual questions about their family's socioeconomic background, self-reported behaviors, and learning and learning contexts, both in the school setting as well as more generally. In compliance with legislation, student questionnaires do not include items about family or personal beliefs (e.g. religious or political beliefs). The student questionnaires focus only on contextual factors that clearly relate to academic achievement.

Educators, psychologists, economists, and others have called for the collection of non-cognitive student information that can explain why some students do better in school than others. Similar questions have been included in other NCES administered assessments such as the Trends in International Mathematics and Science Study (TIMSS), the Program for International Student Assessment (PISA), the National School Climate Survey, and other Federal questionnaires, including the U.S. Census. The insights achieved by the use of these well-established survey questions will help educators, policy makers, and other stakeholders make better informed decisions about how best to help students develop the knowledge and skills they need to succeed.

All questions proposed for piloting have gone through multiple rounds of reviews, including but not limited to reviews by NAEP subject-matter expert groups, organizational Internal Review Board (IRB), and the Governing Board, and have successfully passed extensive pre-testing via cognitive interviews with all respondent groups. Furthermore, NAEP does not report student responses at the individual or school level, but strictly in aggregate forms. To reduce the impact of any individual question on NAEP reporting, the program has shifted to a balanced reporting approach that includes multi-item indices, where possible, to maximize robustness and validity. In compliance with legislation and established practices through previous NAEP administrations, students may skip any question.

A.12. Estimation of Respondent Reporting Burden (2017–2019)

The burden numbers for NAEP data collections fluctuate considerably, with the number of students sampled every other year being much larger than in the years in between. As such, the average annual burden estimates for the three years described in this submission differ from those estimated for any given year.

Exhibit 1 provides the burden information per respondent group, by grade and by year, for the 2017–2019 data collections. Exhibit 2 summarizes the burden across the three years.

A description of the respondents or study is provided below, as supporting information for Exhibit 1:

- **Students** – Students in fourth, eighth, and twelfth grades complete assessment forms that contain 50 or 60 minutes of cognitive blocks¹⁶, followed by non-cognitive block(s) which require a total of 15 minutes to complete. The core non-cognitive items are answered by students across subject areas and are related to demographic information. In addition, students answer subject-specific non-cognitive items. Based on timing data collected from cognitive interviews and previous DBAs, 4th grade students can respond to approximately four non-cognitive items per minute, while 8th and 12th grade students can respond to approximately six non-cognitive items per minute. Using this information, the non-cognitive blocks are assembled so that most students can complete all items in the allocated amount of time. Each cognitive and non-cognitive block is timed so that the burden listed above is the maximum burden time for each student. The administrators and/or test delivery system will move students to the next section once the maximum amount of time is reached. Additional student burden accounts for time to read directions, distribute test booklets (for PBAs), and log on to the computer and view a tutorial (for DBAs). This additional burden is estimated at 10 minutes for PBAs and 15 minutes for DBAs. Therefore, the total burden for students is 25 minutes for PBAs and 30 minutes for DBAs.
- **Teachers** – The teachers of fourth- and eighth-grade students participating in NAEP are asked to complete questionnaires about their teaching background, education, training, and classroom organization. Average fourth-grade teacher burden is estimated to be 30 minutes because fourth-grade teachers often have multiple subject-specific sections to complete. Average eighth-grade teacher burden is 20 minutes if only one subject is taught and an additional 10 minutes for each additional subject taught. Based on timing data collected from cognitive interviews, adults can respond to approximately six non-cognitive items per minute. Using this information, the teacher questionnaires are assembled so that most teachers can complete the questionnaire in the estimated amount of time. For adult respondents, the burden listed is the estimated average burden.
- **Principals/Administrators** – The school administrators in the sampled schools are asked to complete a questionnaire. The core items are designed to measure school characteristics and policies that research has shown are highly correlated with student achievement. A section with subject-specific items concentrates on curriculum and instructional services issues. The burden for school administrators is determined in the same manner as burden for teachers (see above) and is estimated to average 30 minutes per principal/administrator.
- **SD and ELL** – SD and ELL information is provided by school personnel concerning students identified as SD or ELL. This information will be used to determine the appropriate accommodations for students. The burden for school administrators is estimated at 10 minutes, on average, for each student identified as SD and/or ELL.
- **Submission of Samples** – Survey sample information is collected from schools in the form of lists of potential students who may participate in NAEP. This sample information can be gathered manually or electronically at the school, district, or state level. If done at the state level, some states require a data security agreement, which is customized based on the specific requests of the state (see Appendix B for a sample data security agreement). If done at the school or district level, some burden will be incurred by school personnel. It is estimated that it will take two hours, on average, for school personnel to

¹⁶ The assessments given in Puerto Rico are translated into Spanish. To account for the language complexities, additional time is provided for the cognitive blocks (for a total of 80 minutes).

complete the submission process. Based on recent experience, the estimated percent of the schools or districts that will complete the sample submission process depends upon the nature of the sample (i.e., national or state). As such, it is estimated that 19% of the schools or districts will complete the submission process in state assessment years (i.e., 2017 and 2019; based on the data from 2015) and 42% of the schools or districts will complete the submission process in national-only assessment years (i.e., 2018; based on the data from 2014).

- **Pre-Assessment and Assessment Activities** – Each school participating in NAEP has a designated staff member to serve as its NAEP school coordinator. Pre-assessment and assessment activities include functions such as finalizing student samples, verifying student demographics, reviewing accommodations, and planning logistics for the assessment. An electronic pre-assessment system (known as MyNAEP) was developed so that school coordinators would provide requested administration information online, including logistical information, updates of student and teacher information, and the completion of inclusion and accommodation information. More information about the school coordinators’ responsibilities is included in Section B.2. Based on information collected from previous years’ use of MyNAEP, it is estimated that it will take three hours, on average, for school personnel to complete these activities, including looking up information to enter into the system. We will continue to use MyNAEP system data to learn more about participant response patterns and use this information to further refine the system to minimize school coordinator burden.
- **School Coordinator Debriefing Interview** – After each assessment, the field staff will meet with the school coordinator for a debriefing interview. The purpose of this interview is to obtain feedback on how well the assessment went in that school, the usefulness of NAEP materials (e.g., publications, letters, etc.), preparation activities, strategies utilized for increasing participation, and any issues that were noted. A sample of the debriefing interview questions is included in Appendix E-1. It is estimated that this interview will take on average 7 minutes.
- **Post-assessment Follow-up Survey** – As part of the on-going quality control of the assessment process, 25 percent of the schools will be randomly selected for an additional follow-up survey. Survey questions solicit feedback on pre-assessment, assessment, and procedural processes. A sample of a post-assessment follow-up survey is included in Appendix E-2. It is estimated that this interview will take on average 10 minutes.
- **HSTS** – The NAEP HSTS periodically surveys the curricula being followed in our nation’s high schools and the course-taking patterns of high school students through a collection of transcripts. This data collection requires three hours, on average, per school from a sample of approximately 800 schools.
- **NIES** – NIES is designed to describe the condition of education for American Indian and Alaska Native (AI/AN) students in the United States. Additional questionnaires designed for NIES are given to students (estimated at 15 minutes), teachers (20 minutes), and school administrators (30 minutes).
- **CAFS** – The CAFS study contains a supplemental survey questionnaire related to computer familiarity and access. It is given to a subset of students and the time to complete this additional questionnaire is limited to 15 minutes.

EXHIBIT 1

Estimated Burden for NAEP 2017–2019 Assessments, By Year, By Grade Level

(Note: all explanatory notes and footnotes are displayed following the 2019 table)

2017

Subjects	Students			Teachers			School Questionnaire (school principal)			Pre-assessment, sample submission, & assessment feedback (school coordinator)		SD/ELL (school personnel)				Total Burden (in hours)
	# of Students	Avg. minutes per response	Burden (in hours)	# of Teachers	Avg. minutes per response	Burden (in hours)	# of Schools	Avg. minutes per response	Burden (in hours)	# of Schools	Burden (in hours) ¹	# of Schools	# of SD/ELL Students ²	Avg. minutes per response	Burden (in hours)	
4th Grade																
Operational (math, reading, writing, Puerto Rico math); Pilot (reading, math); MST and KaSA studies	349,000	30	174,500	29,702	30	14,851	7,426	30	3,713	7,426	26,274	7,426	80,270	10	13,378	232,716
P&P-DBA Bridge Study (reading, math)	79,000	25	32,917	6,320	30	3,160	1,580	30	790	1,580	5,591	1,580	18,170	10	3,028	45,486
CAFS study ⁵	7,500	15	1,875	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1,875
4th Grade Totals	428,000	N/A	209,292	36,022	N/A	18,011	9,006	N/A	4,503	9,006	31,865	9,006	98,440	N/A	16,406	280,077
8th Grade																
Operational (math, reading, writing, Puerto Rico math); Pilot (reading, math, US history, civics, geography); MST and KaSA studies	372,000	30	186,000	47,489	20 for teachers who teach 1 subject; additional 10 for each additional subject ³	20,262	7,915	30	3,957	7,915	28,006	7,915	66,960	10	11,160	249,385
P&P-DBA Bridge Study (math, reading)	79,000	25	32,917	7,645	20	2,548	1,274	30	637	1,274	4,509	1,274	14,220	10	2,370	42,981
Laptop-DBA Bridge Study (writing)	3,000	30	1,500	720	20	240	120	30	60	120	425	120	540	10	90	2,315
CAFS study ⁵	7,500	15	1,875	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1,875
8th Grade Totals	454,000	N/A	222,292	55,854	N/A	23,050	9,309	N/A	4,654	9,309	32,940	9,309	81,720	N/A	13,620	296,556
Total	882,000	N/A	431,584	91,876	N/A	41,061	18,315	N/A	9,157	18,315	64,805	18,315	180,160	N/A	30,026	576,633

Total number of respondents

1,028,820

Total number of responses

1,205,665

2018

Subjects	Students			Teachers			School Questionnaire (school principal)			Pre-assessment, sample submission, & assessment feedback (school coordinator)		SD/ELL (school personnel)			Total Burden (in hours)	
	# of Students	Avg. minutes per response	Burden (in hours)	# of Teachers	Avg. minutes per response	Burden (in hours)	# of Schools	Avg. minutes per response	Burden (in hours)	# of Schools	Burden (in hours) ¹	# of Schools	# of SD/ELL Students ²	Avg. minutes per response		Burden (in hours)
4th Grade																
Pilot (science)	15,000	30	7,500	1,277	30	639	319	30	160	319	1,276	319	3,450	10	575	10,150
4th Grade Totals	15,000	N/A	7,500	1,277	N/A	639	319	N/A	160	319	1,276	319	3,450	N/A	575	10,150
8th Grade																
Operational (U.S. history, civics, geography, TEL); Pilot (science)	65,000	30	32,500	8,298	20 ³	2,766	1,383	30	691	1,383	5,530	1,383	11,700	10	1,950	43,437
P&P-DBA Bridge Study (U.S. history, civics, geography)	24,000	25	10,000	2,323	20	774	387	30	194	387	1,548	387	4,320	10	720	13,236
8th Grade Totals	89,000	N/A	42,500	10,621	N/A	3,540	1,770	N/A	885	1,770	7,078	1,770	16,020	N/A	2,670	56,673
12th Grade																
Pilot (reading, math, science)	37,000	30	18,500	N/A	N/A	N/A	787	30	394	787	3,148	787	5,550	10	925	22,967
12th Grade Totals	37,000	N/A	18,500	N/A	N/A	N/A	787	N/A	394	787	3,148	787	5,550	N/A	925	22,967
Total	141,000	N/A	68,500	11,898	N/A	4,179	2,876	N/A	1,439	2,876	11,502	2,876	25,020	N/A	4,170	89,790

Total number of respondents

161,527

Total number of responses

183,671

2019

Subjects	Students			Teachers			School Questionnaire (school principal)			Pre-assessment, sample submission, & assessment feedback (school coordinator)		SD/ELL (school personnel)				Total Burden (in hours)
	# of Students	Avg. minutes per response	Burden (in hours)	# of Teachers	Avg. minutes per response	Burden (in hours)	# of Schools	Avg. minutes per response	Burden (in hours)	# of Schools	Burden (in hours) ¹	# of Schools	# of SD/ELL Students ²	Avg. minutes per response	Burden (in hours)	
4th Grade																
Operational (math, reading, science, Puerto Rico math); Pilot (reading, math, writing)	406,500	30	203,250	34,596	30	17,298	8,649	30	4,324	8,649	30,603	8,649	93,495	10	15,583	271,058
NIES ⁴	8,000	15	2,000	2,000	20	667	2,200	30	1,100	N/A	N/A	N/A	N/A	N/A	N/A	3,767
4th Grade Totals	406,500	N/A	205,250	34,596	N/A	17,965	8,649	N/A	5,424	8,649	30,603	8,649	93,495	N/A	15,583	274,825
8th Grade																
Operational (math, reading, science, Puerto Rico math); Pilot (reading, math, writing)	406,500	30	203,250	51,894	20 for teachers who teach 1 subject; additional 10 for each additional subject ³	22,141	8,649	30	4,324	8,649	30,603	8,649	73,170	10	12,195	272,513
NIES	6,500	15	1,625	1,500	20	500	2,000	30	1,000	N/A	N/A	N/A	N/A	N/A	N/A	3,125
8th Grade Totals	406,500	N/A	204,875	51,894	N/A	22,641	8,649	N/A	5,324	8,649	30,603	8,649	73,170	N/A	12,195	275,638
12th Grade																
Operational (math, reading, science); Pilot (reading, math, writing)	70,000	30	35,000	N/A	N/A	N/A	1,489	30	745	1,489	5,270	1,489	10,500	10	1,750	42,765
HSTS	N/A	N/A	N/A	N/A	N/A	N/A	800	180	2,400	N/A	N/A	N/A	N/A	N/A	N/A	2,400
12th Grade Totals	70,000	N/A	35,000	N/A	N/A	N/A	2,289	N/A	3,145	1,489	5,270	1,489	10,500	N/A	1,750	45,165
Total	883,000	N/A	445,125	86,490	N/A	40,606	19,587	N/A	13,893	18,787	66,476	18,787	177,165	N/A	29,528	595,628

Total number of respondents

1,026,652

Total number of responses

1,207,229

Notes for all tables in Exhibit 1

1. The burden for the school coordinator is as follows: Pre-assessment burden is 3 hours, sample submission burden is 2 hours (for 19% of schools in 2017 and 2019 and 42% of schools in 2018, based on 2014 and 2015 data), school coordinator debriefing interview is 7 minutes and post-assessment follow-up survey is 10 minutes (for 25% of the schools).
2. The estimated percent of SD/ELL students (based on the NAEP 2015 sample) is 23%, 18%, and 15%, at grades 4, 8, and 12, respectively.
3. Grade 8 teachers who teach one subject have an estimated burden of 20 minutes, with an additional 10 minutes for each additional subject. There is only one teacher questionnaire for the three social studies subjects (U.S. history, civics, and geography). In 2017 and 2019, the estimated number of teachers who teach 1 subject is 50%, 2 subjects is 45%, 3 subjects is 4%, and 4 subjects is 1%. In 2018, the social studies subjects and TEL will be administered in different schools given that social studies will be administered on tablet and TEL will be administered on laptop. As such, all teachers in 2018 will only receive a questionnaire for one subject area.
4. The burden for NIES is associated with the additional questionnaire that is given to the same students, teachers, and school administrators that respond to the main NAEP questionnaires. As such, the NIES questionnaire does not impact the total number of respondents. The estimated number of students, teachers, and school administrators that will respond to the NIES questionnaires is based on the 2015 sample.
5. The burden for CAFS is associated with the additional questionnaire that is given to the same students that respond to the main NAEP questionnaires. As such, the CAFS questionnaire does not impact the total number of respondents.

EXHIBIT 2

Total Annual Estimated Burden Time Cost for NAEP 2017–2019 Assessments

Data Collection Year	Number of Respondents	Number of Responses	Total Burden (in hours)
2017	1,028,820	1,205,665	576,633
2018	161,527	183,671	89,790
2019	1,026,652	1,207,229	595,628
3-year Annual Average	739,000	865,522	420,684

The estimated respondent burden across all these activities translates into an estimated total burden time cost of \$17,106,770 for 1,262,051 hours¹⁷, broken out by year and respondent group in the table below.

	Students		Teachers and School Staff		Principals		Total	
	Hours	Cost	Hours	Cost	Hours	Cost	Hours	Cost
2017	431,584	\$3,128,984	135,892	\$4,286,034	9,157	\$386,334	576,633	\$7,801,352
2018	68,500	\$496,625	19,851	\$626,101	1,439	\$60,711	89,790	\$1,183,437
2019	445,125	\$3,227,156	136,610	\$4,308,679	13,893	\$586,146	595,628	\$8,121,981
Total	945,209	\$6,852,765	292,353	\$9,220,814	24,489	\$1,033,191	1,262,051	\$17,106,770

A.13. Cost to Respondents

There are no direct costs to respondents.

¹⁷ This is based on 945,209 hours for students at \$7.25 an hour (based on the federal minimum wage), 292,353 hours for teachers and school staff at \$31.54 an hour (based on a 10-month salary from data from Bureau of Labor Statistics, U.S. Department of Labor, The Economics Daily, *Employment and annual wages for preschool, primary, middle, and secondary school teachers*, on the Internet at <http://www.bls.gov/opub/ted/2015/employment-and-annual-wages-for-preschool-primary-middle-and-secondary-school-teachers.htm> [visited December 08, 2015]), and 24,489 hours for principals at \$42.19 an hour (based on data from Bureau of Labor Statistics, U.S. Department of Labor, Occupational Outlook Handbook, 2014-15 Edition, *Elementary, Middle, and High School Principals*, on the Internet at <http://www.bls.gov/ooh/management/elementary-middle-and-high-school-principals.htm> [visited December 8, 2015]).

A.14. Estimates of Cost to the Federal Government

The total cost to the federal government for the administrations of the 2017–2019 activities is estimated to be approximately \$94.6 million for the three years (annualized average of \$31.5 million). The 2017–2019 cost estimate is broken down as follows:

- \$2.2 million for the printing, packaging, and distribution phases of the administrations.
- \$84.8 million for the cost of the field supervisors and data collectors to go into schools to administer the 2017–2019 assessments, including travel expenses and testing equipment costs; and
- \$7.6 million for web operations and maintenance costs related to the support of DBAs.

A.15. Reasons for Changes in Burden (from last Clearance submittal)

The nature of NAEP is that burden alternates from a relatively low burden in national-level administration years (i.e., even years) to a substantial burden increase in state-level administration years that include one or more assessments that support national, state-by-state, and certain urban districts reporting (i.e., odd years). In state/district assessment years, NAEP samples approximately 1,000,000 students, while in national-only assessment years, approximately 100,000 students. In 2017 and 2019, NAEP will conduct state/district assessments, and in 2018, national-level assessments. The previous three-year clearance included burden for one state/district assessments year (2015) and only two national-level assessments years (2014 and 2016), therefore the overall number of respondents and responses is larger in this clearance request than in the previous one.

In addition, recent reports from the field staff have indicated that the pre-assessment activities require three hours, rather than the two hours previously estimated. Therefore, we have adjusted the burden estimate accordingly.

However, because NAEP is seeking a new OMB number at this time, there is no change shown in the OMB system under the new OMB number from the burden approved under NAEP’s previous OMB#.

A.16. Time Schedule for Data Collection and Publications

The time schedule for the data collection for the 2017–2019 assessments is shown below.

2017	January–March 2017
2018	January–March 2018
2019	January–March 2019

The grades 4 and 8 reading and mathematics national and state results are typically released to the public around October of the same year (i.e., about 6-7 months after the end of data collection). However, note that the PBA validation studies planned in 2017 may delay that particular release. All other operational assessments are typically released 12-15 months after the end of data collection.

The operational schedule for the assessments generally follows the same schedule for each assessment cycle. The dates below show the specifics for the 2019 state-level assessments:

- Spring 2018: Select the school sample and notify schools
- October – November 2018: States, districts, or schools submit the list of students
- December 2018: Select the student sample
- December 2018 – January 2019: Schools prepare for the assessments using the MyNAEP system
- January – March 2019: Administer the assessments
- March – May 2019: Process the data, score constructed response items, and calculate sampling weights
- June – July 2019: Analyze the data
- July – September 2019: Prepare the reports, obtaining feedback from reviewers

- October 2019: Release the results

A.17. Approval for Not Displaying OMB Approval Expiration Date

No exception is requested.

A.18. Exceptions to Certification Statement

No exception is requested.