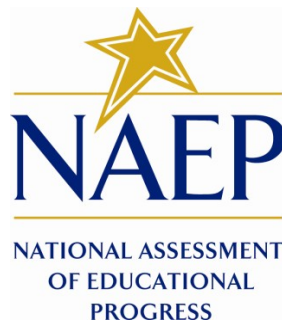


NATIONAL CENTER FOR EDUCATION STATISTICS
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Volume I
Supporting Statement

***National Assessment of Educational Progress (NAEP)
Digitally Based Assessments (DBA) Usability Study 2017-18***

OMB# 1850-0803 v.181



November 2016

Table of Contents

1	SUBMITTAL-RELATED INFORMATION.....	1
2	BACKGROUND AND STUDY RATIONALE.....	1
3	RECRUITMENT AND SAMPLE CHARACTERISTICS.....	2
4	STUDY DESIGN AND DATA COLLECTION.....	2
5	CONSULTATIONS OUTSIDE THE AGENCY.....	4
6	JUSTIFICATION FOR SENSITIVE QUESTIONS.....	4
7	PAYING RESPONDENTS.....	4
8	ASSURANCE OF CONFIDENTIALITY.....	4
9	ESTIMATE OF HOURLY BURDEN.....	4
10	COST TO FEDERAL GOVERNMENT.....	5
11	PROJECT SCHEDULE.....	5

1 SUBMITTAL-RELATED INFORMATION

This material is being submitted under the generic National Center for Education Statistics (NCES) clearance agreement (OMB# 1850-0803), which provides for NCES to conduct various procedures (such as pilot tests, cognitive interviews, and usability studies) to test new methodologies, question types, or delivery methods to improve survey and assessment instruments.

NAEP is planning to conduct digitally based assessment (DBA) usability studies in 2017 and 2018, as has been done in the past few years. The usability testing activities anticipated for 2017 and 2018 will follow the same protocol as the last approved NAEP DBA Tools and Item Types Usability Study 2015-16 (OMB# 1850-0803 v.142), with a few changes in regard to the project timeline and recruitment plan:

1. The recruitment plan no longer enlists the help of NAEP State Coordinators (NSC). Instead, all recruitment will be done through EurekaFacts. The burden table and cost estimates have been updated to reflect this change.
2. This submission covers two years of usability testing activities, whereas the previous package covered only one year. The burden table, cost estimates, and project schedule have been updated to reflect this longer window of usability testing.

2 BACKGROUND AND STUDY RATIONALE

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, and the arts. NAEP is administered by NCES, which is part of the Institute of Education Sciences, within the U.S. Department of Education. NAEP's primary purpose is to assess student achievement in the different subject areas and collect survey questionnaire (i.e., non-cognitive) data to provide context for the reporting and interpretation of assessment results.

NCES is interested in studying the intuitiveness and ease-of-use of interface tools and item types being developed for future NAEP assessments. Examples of recently developed tools include an in-system protractor and an in-system bar graph creator. Tools like these provide support to the student during the assessment and allow greater flexibility in testing cognitive constructs. As part of the process of designing new tools and item types, it is important to study the user experience they provide. User testing with students is crucial for making design decisions that result in tools that are truly useful and intuitive to use. User testing can reveal usability problems that were unanticipated by developers and designers. In comparison to previous usability studies¹, this study will include a greater emphasis on new ways of interacting with items and new item types, as well as a tutorial.

This study will present students with prototypes of new item types, tools, and a tutorial on touch-screen tablets. Students will be given assessment-related tasks to complete using the prototypes (see Volume II). Results will consist of task completion success rates using each of the control and information elements studied. These task completion success rates will be combined with qualitative information from ease-of-use surveys, exit questions, and facilitator observations to compile recommendations for modifications to items and tools. This is done to ensure that their implementation in upcoming assessments does not present barriers to construct validity. In addition, some of the information will be used to assess the usability of the hardware being used in the study. Screen size, keyboard, and trackpad responsiveness are just some of the hardware properties that affect usability.

¹ NCES has conducted usability studies for DBA development over the past few years, as described in OMB# 1850-0803 v.112 and 1850-0803 v.142. This submission is based on the last approved DBA usability test submission (OMB# 1850-0803 v.142).

Information on ease-of-use of the hardware will be used to inform future decisions regarding selection of appropriate systems for NAEP testing.

Volume I of this submittal describes the purpose, design, sampling, burden, cost, and schedule information for the study. Volume II provides examples of the types of tasks that will be included in the protocol as well as the types of survey questions that will be administered in the study. The appendices contain recruitment materials, notifications, usability testing scripts, and thank you documents.

3 RECRUITMENT AND SAMPLE CHARACTERISTICS

Fulcrum, the NAEP contractor for development and support of NAEP digitally based assessments (DBA) for NCES will manage the usability study. Fulcrum will utilize EurekaFacts (see section 5) as a contractor to provide participants on an as-needed basis. EurekaFacts will recruit participants from the greater Washington, DC/Baltimore metropolitan area using various outreach methods. These methods will include over-the-phone recruitment based on targeted mailing lists containing residential addresses and landline telephone listings, newspaper/internet ads, outreach to community organizations (e.g., Boys and Girls Clubs, Parent-Teacher Associations), and mass media (e.g., postings on the EurekaFacts website). By using EurekaFacts for recruitment, Fulcrum will be able to administer usability testing during school holidays, evenings, and weekends allowing for the total number of participants to be spread out over the course of the calendar year, rather than being confined to the school calendar.

When recruiting participants, EurekaFacts staff will first communicate with the parent/guardian of the interested minor. The parent/guardian will be informed of the objectives, purpose, and participation requirements of the study and the activities it entails. After confirming that a student is to participate, a confirmation e-mail/letter will be sent and the informed parental/guardian consent for the minor's participation will be obtained. The appendices provide sample recruitment materials that will be used by EurekaFacts.²

While no demographic variables have been shown to affect outcomes in past usability studies, recruiters will make an effort to recruit a diverse sample of students in order to minimize systematic variance in the study sample. Though we strive to ensure that participants are as diverse as is practical, students chosen for the study will not be included or excluded based on demographic criteria. Given the potentially large number of interactions to be tested, up to 200 students will participate in user testing each year, spread across grades 4, 8, and 12³. As such, a total of up to 400 students may participate over the two year period covered by this submission.

4 STUDY DESIGN AND DATA COLLECTION

Recruited students will participate in the usability test at EurekaFacts or another suitable venue (e.g., a school library or after-school office) outside of school hours. In order to ensure consistency and internal validity predetermined testing scripts, materials, and experimental stimuli will be used.

User testing will be conducted in several sessions over the course of the year, as part of an iterative process of design and testing of new and revised DBA tools, tutorial, and items developed over that period. DBA developers and designers will submit prototypes designed to test specific interactions, and these prototypes will be used in subsequent user testing groups. User testing data will be reported back to the developers and NCES as they are

² Note: If appropriate, relevant appendices (e.g., parental screening calls) may be translated to another language to facilitate communication.

³ If sessions are conducted in the summer, students may have just completed the target grade or be entering into the target grade.

collected so that decisions regarding design modifications can be made in a timely manner. Modified features or items may then be included in a later user testing session to validate the usability of the changes.

A variety of subject areas will be included, not to test the subject content, but to test interactions that may be unique to items for that subject. For example, math items may be used to test an on-screen calculator or equation editor, as that subject area uses those two particular interactions. Reading items may be used to test different passage layouts and panel controls that are unique to reading items.

In addition to the multiple item types tested using prototypes, different participant groups may be tested using different touch-screen tablets, in order to test the impact of different hardware or operating systems on the usability and the interactions.

Each student will perform the study tasks during a one-on-one session with a facilitator. For some of the tasks, the facilitator will give verbal instructions, such as *“Imagine that you want to change the color of the tool bar up there [point] from black to white. Please show me how you would do that.”* For other tasks, students will be instructed to follow the written instructions on the screen or to attend a tutorial. For most tasks, participants will be asked to explain what they are doing and why, as they perform the tasks.

User testing will take no more than 75 minutes per student. Students will be allowed to take breaks as needed. Screen capture software for user testing, such as Morae, may be used as appropriate to document on-screen activity for later analysis. Depending on the needs of the analysis, keyboard logging and clickstream recording may also be performed. Screen activity may be recorded but the participants themselves will not be recorded.

Students’ success or difficulty in completing assigned tasks will be analyzed to determine which information or control elements are missing or insufficient to allow successful completion of anticipated user tasks. While successful completion of tasks will be recorded, it is the tools and item types that are being evaluated rather than the students. All results will be used only to make recommendations regarding the design and development of tools and item types.

Results will be analyzed chiefly in terms of descriptive statistics detailing the distribution of success rates and subjective user ratings. An example finding would be: “40 percent of participants found the volume control without assistance.” This finding would then be used to determine that the volume control needs to be made more visible to users in order to be used successfully by 100 percent of the students. Other statistical comparisons may be performed as appropriate to the variables and populations.

The following instruments will be used to gather data, and are included in the Volume II protocol.

1. *Participant ID and Welcome Script* – This welcome script introduces the interviewer and explains the study.
2. *Computer and Tablet Familiarity Survey* – This survey will be used to determine if differences in student performance can be attributed to previous experiences using computers or touch-screen tablets.
3. *Ease-of-Use Rating Survey* – This survey, completed by the facilitator, is used to record ease-of-use ratings for tasks.
4. *User Testing Scenarios* – This protocol contains the script for the facilitators to guide the interactions of the participants with the tablets. Sample tasks are included in Volume II to give reviewers a sense of what will be tested. The specific testing scenarios will be determined throughout the system development. Data will be recorded on a tablet or laptop used by the facilitator.
5. *Exit Questions* – These sample questions can be asked at the completion of the user tasks.

5 CONSULTATIONS OUTSIDE THE AGENCY

Fulcrum IT Services, LLC is the NAEP contractor for the development and ongoing support of NAEP digitally based assessments for NCES, including the system used for this usability study of tools and item types. Fulcrum will carry out the planning and analysis and will produce the final report.

Educational Testing Service (ETS) serves as the Item Development (ID) and Design, Analysis, and Reporting (DAR) contractor on the NAEP project, developing cognitive and survey items for NAEP assessments. ETS staff will be involved in item development activities and may assist in administering and/or observing some of the user testing sessions.

EurekaFacts is a research and consulting firm in Rockville, Maryland, that offers facilities, tools, and staff to collect and analyze both qualitative and quantitative data. EurekaFacts will be involved in recruitment, logistics, and administering the studies with participants.

6 JUSTIFICATION FOR SENSITIVE QUESTIONS

Throughout the item and protocol development processes, effort has been made to avoid asking for information that might be considered sensitive or offensive.

7 PAYING RESPONDENTS

To encourage participation and thank students for their time and effort, a \$25 gift card will be offered to each participating student. If a parent or legal guardian brings their child to and from the testing site, they will also receive a \$25 gift card along with a thank you letter for allowing their child to participate in the study (Appendix L).

8 ASSURANCE OF CONFIDENTIALITY

The usability study will not retain any personally identifiable information. Prior to the start of the usability study, participants will be notified that their participation is voluntary and that their answers may be used only for research purposes and may not be disclosed, or used, in identifiable form for any other purpose except as required by law [Education Sciences Reform Act of 2002 (20 U.S.C. §9573)].

Written notification will be sent to legal guardians of students before user testing is conducted. Participants will be assigned a unique identifier (ID), which will be created solely for data file management and used to keep all participant materials together. The participant ID will not be linked to the participant name in any way or form.

Screen actions with audio may be captured from each session. The only identification included on the files will be the participant ID. The screen capture will be used for analysis after the session. Small portions of the screen capture for select sessions may be used in NCES briefings in order to demonstrate the methodology used for this study. No personally identifying information will be included in data analyses or study briefings.

9 ESTIMATE OF HOURLY BURDEN

The estimated burden for recruitment assumes attrition throughout the process. Assumptions for approximate attrition rates for direct participant recruitment from initial contact to follow-up are 40%, and from follow-up to confirmation are 50%.

All sessions will be scheduled for no more than 75 minutes. Estimated hourly burden for the participants is

described in Table 1, below. Participants will be subsets of the initial contact group.

Table 1. Estimate of Hourly Burden

Respondent	Hours per respondent	Number of respondents	Number of responses	Total hours (rounded up)
Schools and Organizations				
Initial contact	0.05	200	200	10
Follow-up contact	0.15	120*	120	18
Confirmation	0.05	60*	60	3
Sub-Total		200	380	31
Parent or Legal Guardian for Student Recruitment				
Initial contact	0.05	1,468	1,468	74
Follow-up contact	0.15	880*	880	132
Consent form completion and return	0.13	440*	440	58
Confirmation	0.05	440*	440	22
Sub-Total		1,468	3,228	286
Participation (User Testing)				
Students	1.25	400 ^a	400	500
Sub-Total		400	400	500
Total Burden		2,068	4,008	817

* Subset of initial contact group, not double counted in the total number of respondents.

^a Estimated number of actual participants will be somewhat less than confirmation numbers.

10 COST TO FEDERAL GOVERNMENT

Table 2 provides the overall project cost estimates.

Table 2: Estimate of Costs

Activity	Estimated Cost
Fulcrum Study Design and Analysis	\$38,332
EurekaFacts Costs for Student Recruitment and Administration of User Testing	\$562,400
Total	\$600,732

11 PROJECT SCHEDULE

Table 3 provides the overall schedule.

Table 3: Schedule

Event	Date
Ongoing Recruiting for User Testing Sessions (subsequent to OMB Clearance)	December 2016 – July 2018
User Testing	January 2017 – August 2018
Summary Report	September 2017 and September 2018