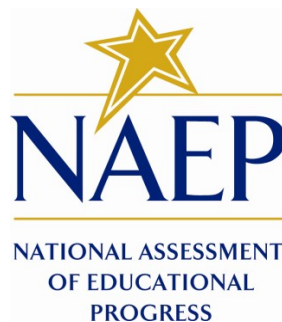


NATIONAL CENTER FOR EDUCATION STATISTICS
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Volume I
Supporting Statement

NAEP Digitally Based Assessments (DBA)
Tools and Item Types Usability Study 2015-16

OMB# 1850-0803 v.142



August 10, 2015

Contents

1	SUBMITTAL-RELATED INFORMATION.....	1
2	BACKGROUND AND STUDY RATIONALE.....	1
3	RECRUITMENT AND SAMPLE CHARACTERISTICS.....	1
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.....	5
9	ESTIMATE OF HOURLY BURDEN.....	5
10	COST TO FEDERAL GOVERNMENT.....	7
11	PROJECT SCHEDULE.....	7

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.

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 for 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 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 studies, this study will include a greater emphasis on new ways of interacting with items and new item types. In addition, the 2017 eNAEP assessment will involve more subject areas than previous years, and thus all previous mechanisms will also need to be assessed for the appropriateness of content.

This study will present students with prototypes of new item types and tools 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 and keyboard and trackpad responsiveness are just some of the hardware properties that affect usability. 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 type of tasks that will be included in the protocol as well as types of survey questions that will be administered in the study. The appendices contain sample 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. Two recruitment methods will be used. In one method, NAEP State Coordinators (see section 5) will leverage their relationships within the states to recruit public schools for the usability study.

These schools will then be contacted by Fulcrum for on-site coordination of the study at the school. In a similar manner, NCES representatives who work with private schools and the Bureau of Indian Affairs will leverage their relationships with schools for this study.

Recruiting activities may include presentations about the study at NCES-sponsored workshops and meetings. Flyers will be handed out to interested parties at NCES-sponsored events (Appendix A). Schools may participate in more than one user testing session if they wish. However, individual students may participate in only one session in order to prevent learning effects and to reduce individual burden. Prior to user testing at schools, Fulcrum staff will contact participating schools to make logistical arrangements. Initial contact will be made via email (Appendix B), with a follow-up call to gather necessary logistical information (Appendix C). A parental notification letter (Appendix D) will be provided to the schools for their use to notify parents or guardians of students in the study.

For the second recruitment method, 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 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 (such as postings on the EurekaFacts website). By using EurekaFacts for a portion of the recruiting, Fulcrum will be able to conduct studies during school holidays, 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 about 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. Appendices F through W provide sample recruitment materials that will be used by EurekaFacts.¹

We anticipate that approximately 70% of study participants will be obtained via schools and state coordinators, and 30% through the use of EurekaFacts. Students who have already participated in studies through their school will not be recruited via EurekaFacts, and vice-versa, for reasons of internal study validity and public burden.

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 it 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 360 students will participate in the user testing spread across grades 4, 8, and 12.

4 STUDY DESIGN AND DATA COLLECTION

In schools recruited by the NAEP state coordinators, Fulcrum will bring all necessary materials, including computers and tablets, on the day of the study. Fulcrum will work with school administrators to minimize academic disruption to student participants, and will develop study schedules with that goal in mind and based on the needs of the schools. We anticipate testing approximately 18 students from each participating school.

¹ Note: If appropriate, relevant appendices (e.g., parental screening calls) may be translated to another language to facilitate communication.

Students recruited through EurekaFacts will participate in cognitive interviews at EurekaFacts or another suitable venue (e.g., a school library or after-school office). In order to ensure consistency and internal validity, Fulcrum will facilitate the study and bring all necessary materials, including computers and tablets, to the study location.

User testing will be conducted in small batches over the course of the school year, as part of an iterative process of design and testing of new and revised DBA tools 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 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.

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 & 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 usability study and produce the final report.

The NAEP State Coordinators serve as liaisons between their state education agency and NAEP, coordinating NAEP activities in their states. For usability studies, their chief role is to contact schools to secure participation. Interested schools are then passed to Fulcrum for logistical arrangement of the study.

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 conducting 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 and logistics on an as-needed basis.

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

For sessions recruited by state coordinators: Student participants will not receive financial incentives for participation in this study. However, students will be permitted to keep the earbuds they used during the study. Schools participating in the usability study will each be offered a \$50 gift card to a major office/school supply store as a token of appreciation. Gift cards will be provided at the end of the site visit, and a thank you email will be sent that evening to the school contact (Appendix E).

For sessions recruited by EurekaFacts: To encourage participation and thank students for their time and effort, a \$25 credit card 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 Q). Students will also be able to keep the earbuds used in the study.

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

For the sessions recruited by the state coordinators: School administrators will provide pre-assessment information and help with the logistics of student and room coordination and other related duties. The school administrator burden is estimated at 12 minutes for the pre-assessment contact and up to one hour for administration support. It is possible that up to 26 schools will be contacted initially in order to identify the approximately 14 schools required for the study.

Parents of participating students will receive a letter explaining the study, for which the parent’s burden is estimated at three minutes. An additional burden (15 minutes) is estimated for a small portion of parents (10%) who may write refusing approval for their child or may research information related to the study.

For sessions recruited by EurekaFacts: 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 50%.

All sessions will be scheduled for no more than 75 minutes. For the purposes of estimating burden, it is assumed that 70% of the students will be recruited by the state coordinators and 30% of the students will be recruited by EurekaFacts. Estimated hourly burden for the participants is described in Tables 1, 2, and 3, below. Participants will be subsets of the initial contact group.

Table 1. Estimate of Hourly Burden Using State Coordinator Recruitment (estimated at 70% of participants)

Respondent	Task	Hours per respondent	Number of respondents	Number of responses	Total hours
School Administrator	Initial Contact by NAEP State Coordinator	0.2	26	26	6
School Personnel	Scheduling & Logistics	1	14	14	14
Parents/Legal Guardians	Initial Notification	0.05	280	280	14
Parents/Legal Guardians*	Researching Study	0.25	28*	28	7
Students	User Testing	1.25	252	252	315
Total Burden			572	600	356

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

Table 2. Estimate of Hourly Burden Using EurekaFacts Recruitment (estimated at 30% of participants)

Respondent	Hours per respondent	Number of respondents	Number of responses	Total hours
Schools and Organizations				
Initial contact	0.05	100	100	5
Follow-up contact	0.15	60*	60	9
Confirmation	0.05	30*	30	2
Sub-Total		100	190	16
Parent or Legal Guardian for Student Recruitment				
Initial contact	0.05	400	400	20
Follow-up contact	0.15	240*	240	36
Consent form completion and return	0.13	120*	120	16
Confirmation	0.05	120*	120	6
Sub-Total		400	880	78
Participation (User Testing)				
Students	1.25	108 ^a	108	135
Sub-Total		108	108	135
Total Burden		608	1,178	229

* 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.

Table 3. Total Estimate of Hourly Burden

Recruitment Method	Number of respondents	Number of responses	Total hours
State Coordinators	572	600	356
EurekaFacts	608	1,178	229
Total Burden	1,180	1,778	585

10 COST TO FEDERAL GOVERNMENT

Table 4 provides the overall project cost estimates.

Table 4: Estimate of Costs

Activity	Estimated Cost
Fulcrum Costs for School Recruitment, User Testing, Analysis	\$171,345
EurekaFacts Costs for Student Recruitment	\$52,604
Total	\$223,949

11 PROJECT SCHEDULE

Table 5 provides the overall schedule.

Table 5: Schedule

Event	Date
Ongoing Recruiting for User Testing Sessions (subsequent to OMB Clearance)	August 2015 - September 2016
User Testing	September 2015 - July 2016
Summary Report	August 2016