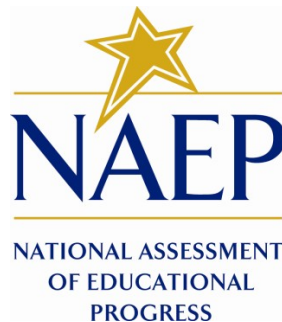NATIONAL CENTER FOR EDUCATION STATISTICS
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Volume I
Supporting Statement

Usability Study for Use of Touch-Screen Tablets

in NAEP Assessments

OMB# 1850-0803 v.87

Contents

NAEP Tablet Usability Study

1 SUBMITTAL-RELATED INFORMATION

This material is being submitted under the generic National Center for Education Statistics (NCES) clearance agreement (OMB #1850-0803) that provides for NCES to conduct various procedures (such as field tests, cognitive interviews, 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, part of the Institute for Education Sciences, in the U.S. Department of Education. NAEP's primary purpose is to assess student achievement in the various subject areas and to also 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 feasibility of using touch-screen tablets for delivery of future NAEP assessments. The use of tablets is perceived to have the following benefits:

1. Flexibility –Tablets can be used for an entire school day without needing to be plugged in or recharged. This allows assessments to be conducted in a wider range of rooms than with laptop computers.
2. Portability – Tablets are generally lighter than laptops. This makes setting up an assessment lab in a school much easier for the assessment personnel and reduces the amount of time the assessment personnel must utilize the classroom.
3. Familiarity – Touch-screens are increasingly common on personal devices, and may be more familiar to some students than traditional laptops. In an effort to make the technology of the NAEP assessments deviate as little as possible from the systems students use most frequently, tablets may be the best option as smart phones and tablets increase in usage among both children and adults.
4. Accessibility – Touch-screen tablets have multiple accessibility features built into their operating systems. They can zoom with the touch of the hand, do not require a mouse, and can be easily adjusted from vertical to horizontal orientation. This flexibility may make it easier for NAEP to accommodate a greater number of students in its assessments.

Many of the attributes that make tablets attractive for administering NAEP assessments are mechanisms that differ substantially from those available to students on a laptop computer. As a result, it is important to study the impact of the changes in screen size and input device (touch screen vs. mouse) on students' ability to interact with assessment items.

This study will present students with item prototypes (see Section 4) on a tablet. Students will be given assessment-related tasks to complete using the tablet (see Volume 2). Results will consist of student success rates using each of the control and information elements tested, along with recommendations for modifications to items and the assessment environment (e.g., system tools, hardware) to ensure highly successful usability during the upcoming assessments.

Volume I of this submittal contains descriptions as well as design, sampling, burden, cost, and schedule information for the study. Volume II contains examples of the type of tasks that will be included in the protocol as

well as survey questions of the type that will be administered in the study. The appendices contain sample scripts, notifications, and thank you documents.

3 RECRUITMENT AND SAMPLE CHARACTERISTICS

NAEP State Coordinators (see section 5) will leverage their relationships within the states to contact schools and identify those schools willing to participate in the usability study. Only schools that have not been selected to participate in NAEP assessments in the 2013-2014 school-year will be selected for this study so that students will not have had recent prior experience with NAEP assessments. Schools may participate in more than one user testing session if they wish, though students may participate in only one session, in order to prevent learning effects and to reduce individual burden. The NAEP State Coordinators will forward the contact information for participating schools to Fulcrum IT, the NAEP web technology contractor (see section 5). Fulcrum IT will contact each school to make the arrangements for students from that school to participate.

Given the number of interactions to be tested, up to 280 students will participate in the user testing, spread approximately equally from grades 4, 8, and 12. These students will be selected from 20-40 schools, allowing for a range of participants from rural and urban areas as well as different SES levels. Students chosen for the study will not be included or excluded based on these demographic criteria. However, a computer and tablet familiarity survey (see Volume II) will be administered to participants during the study to help determine if pre-study familiarity with tablets has an impact on the results of this usability study.

4 STUDY DESIGN AND DATA COLLECTION

Prior to the user testing, Fulcrum IT staff will contact cooperating schools to make logistical arrangements. Initial contact will be made via email (see appendix A), with a follow-up call to gather necessary logistical information (see appendix B). An alternate follow-up script will be used for subsequent site visits to a particular school (see appendix C). A parental notification letter (see appendix D) will be provided to the schools for their use to notify parents or guardians of students in the study. Fulcrum will bring all necessary materials, including tablets, to the schools on the day of the study.

User testing will be conducted in small batches over the course of 9 months, as part of an iterative process of design and testing of new and revised interactions developed over that period for tablet-based assessments.

Developers will submit prototypes designed to test specific interactions with tablet features or items, and these prototypes will be used in a subsequent user testing group. User testing data will be reported back to the developers as they are collected so that they can make any modifications necessary to improve the usability of the system or items. 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 will be used to test an on-screen calculator or equation editor, as that is the only subject area that uses those two particular interactions. The study will also include reading items in order to test certain layouts that are used only for items requiring intuitive, simultaneous access to multiple prompts and response 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 of 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 [points] 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 to 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. Any interactions that cannot be tested in the allotted time will be moved to a later user testing session with different participants. Students will be allowed to take breaks as needed. Each user testing session will be recorded for later analysis. The recording will capture everything that happens on the computer screen, as well as audio of the facilitator and the participant. No video of the students or their surroundings will 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 tablet hardware and operating system, and the item interactions that are being evaluated, rather than the students, and all results will be used to make recommendations regarding administering assessments using tablets.

Results will be analyzed in terms of the percent of participants who successfully completed each task and task element. An example finding would be: "40% of participants found the volume control without assistance." This finding would be used to determine that the volume control needed to be made more visible to users in order to be used successfully 100% of the time.

In addition, ease of use ratings will be collected for the tasks, as well as any comments the participants have. The following instruments that will be used to gather data are included in the Volume II protocol.

1. *Demographic Information and Welcome Script* – The demographic table will be filled out by the facilitator with information about each participant. The 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 attributable to previous experience using touch-screen tablets.
3. *User Testing Scenarios* – This protocol contains the script for the facilitators to guide the interactions of the participants with the tablets. General navigation and display mechanisms, as well as a subset of items related to reading assessments, are described. Data will be recorded in a spreadsheet on a tablet or laptop used by the facilitator.
4. *Ease-of-Use Rating Survey* – This survey, completed by the facilitator, is used to record ease-of-use ratings for tasks.
5. *System Usability Scale (SUS)* – This is a standardized norm-referenced tool that will be administered at the end of each session by the facilitator. While not all questions are relevant to this study, the entire tool will be administered in order to comply with the standardized administration methods of the tool. The SUS is used

solely to gauge usability among different systems, and will not be used to make recommendations for tablet use or item development.

5 CONSULTATIONS OUTSIDE THE AGENCY

Fulcrum IT LLC (Fulcrum IT)

Fulcrum IT is the NAEP contractor responsible for the development and ongoing support of NAEP computer-based assessments for NCES, including the system to be used for the touch-screen tablet usability study. Fulcrum IT will conduct the usability study and produce the final report.

NAEP State Coordinators

The NAEP State Coordinator serves as the liaison between the state education agency and NAEP, coordinating NAEP activities in his or her state. As previously noted, NAEP State coordinators from selected states will work with schools within their states to identify participating schools.

Educational Testing Service (ETS)

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 some item development activities and may assist in conducting and/or observing some user testing sessions.

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

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 (see appendix E).

8 ASSURANCE OF CONFIDENTIALITY

The usability study will not collect 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 statistical 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.

The user testing sessions will be recorded. The only identification included on the files will be the participant ID. In addition, the researchers may wish to use some recordings in scientific presentations as a means of describing the methodology used for this study. Any recordings used would not be paired with the participant's name, address, or likeness.

9 ESTIMATE OF HOURLY BURDEN

School administrators 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 50 schools will be contacted initially in order to identify the 20-40 required for the study.

Parents of participating students will receive a letter explaining the study, for which the parent's burden is estimated at 3 minutes. An additional burden (15 minutes) is estimated for a small portion of parents (up to 40) who may write refusing approval for their child or may research information related to the study.   As such, it is possible that up to 320 parents will need to be notified in order to identify the 280 students required for the study. 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

Person	Task	Number of Individuals	Individual Participant Burden	Total Burden
School Administrator	Initial Contact by NAEP State Coordinator	50	12 minutes	10 hours
School Personnel	Scheduling & Logistics	40	60 minutes	40 hours
Parents	Initial Notification	320	3 minutes	16 hours
Parents*	Researching Study	40*	15 minutes	10 hours
Students	User Testing	280	75 minutes	350 hours
Total (730 responses)		690	NA	426 hours

* These parents are a subset of those who were initially notified.

COST TO FEDERAL GOVERNMENT

Table 2, below, provides the overall project cost estimates.

Table 2: Estimate of Costs

Activity	Provider	Estimated Cost
Recruitment, User Testing, Analysis	Fulcrum IT	$430,347
Incentive payments	Fulcrum IT	$2,000
Total		$432,347

10 PROJECT SCHEDULE

Table 3, below, provides the overall schedule.

Table 3: Schedule

Date	Event
November 2013*-August 2014	Recruiting for User Testing Sessions
December 2013-August 2014	User Testing
September 2014	Summary Report

*Recruiting will begin immediately upon receipt of OMB Clearance.