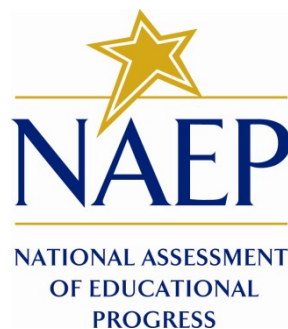


NATIONAL CENTER FOR EDUCATION STATISTICS  
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

*Volume I*  
*Supporting Statement*

*NAEP Survey Assessments Innovations Lab (SAIL)*  
*Science Projects: Interactive Virtual Models*  
*and Virtual Science Lab*  
*Playtesting and Cognitive Interviews*

OMB# 1850-0803 v.151



February 2016  
Revision to a previously approved package (1850-0803 v.100)

## Table of Contents

1)	Submittal-Related Information.....	2
2)	Background and Study Rationale.....	2
3)	Sampling and Recruitment Plans.....	4
4)	Data Collection Process.....	5
5)	Consultations Outside the Agency.....	8
6)	Assurance of Confidentiality.....	8
7)	Justification for Sensitive Questions.....	8
8)	Estimate of Hourly Burden.....	8
9)	Estimate of Costs for Recruiting and Paying Respondents.....	10
10)	Costs to Federal Government .....	10
11)	Schedule.....	11

## **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 field tests and cognitive interviews) to test new methodologies, question types, or delivery methods to improve assessment instruments.

This package is a revision to a previously approved package for the NAEP Survey Assessments Innovations Lab (SAIL) Science Projects: Interactive Virtual Models and Virtual Science Lab (1850-0803 v.100, NAEP SAIL Science Development 2014-15). Given that the development of the materials to be tested has taken longer than anticipated, none of the pretesting work described in the package has yet been conducted. In addition, as the development of these projects has progressed, the pretesting needs have been adjusted to better fit the needs of these innovative studies. In particular, the researchers have determined that student feedback on the tasks and being able to understand students' actions and comprehension of the tasks is critical to inform development. Given that tryouts do not provide this feedback, they have been removed for both projects. Playtesting provides some feedback from students, but not at the level necessary for the Virtual Models project and, as such, playtesting has been removed for the Virtual Models project. In addition, the sample size has been modified for the remaining playtesting and cognitive interview studies, and a second setting (or version) in which to conduct the Virtual Models study has been added. Finally, the timeframe to conduct the pretesting has been adjusted to reflect delays in the development, as well as the revised pretesting plans. This submission is a revision to the originally approved package (1850-0803 v.100) reflecting these updated plan.

## **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.

As part of NAEP's development process, systems of delivery and assessment items are pretested on smaller numbers of respondents before they are administered to larger samples in pilot or operational administrations. The NAEP Survey Assessments Innovations Lab (SAIL) initiative is a research program set up to explore the potential value to NAEP of developing and administering innovative technology-based item types. The current projects seeking approval for conducting empirical studies are two SAIL science-related projects: SAIL Interactive Virtual Models and SAIL Virtual Science Lab.

Both of these projects have been proposed to develop systems or tools that can be used as a platform for assessment activities or tasks. Interactive Virtual Models are envisaged as screen-based representations of real science phenomena and systems that students can interact with to show their knowledge of or reason about those phenomena; the two versions in the present research project will be an interactive model of our solar system (earth, moon, sun) and a hybrid hands-on categorization task (inspecting natural objects

and placing them into categories based on observation of their physical characteristics). The Virtual Science Lab is envisaged as a tablet-based interactive science laboratory in which students can operate virtual science tools, run a wide range of virtual experiments, and collect scientific data; the experimental variables and virtual equipment in the lab will have realistic operational qualities. The studies proposed here are designed to gather information about how these platforms are used by students, the kinds of activities that we can have students do with them, and the kinds of information they can provide for assessment purposes. The information gathered from the proposed studies will feed into the iterative development of these interactive systems.

As part of the SAIL research and development process, the interactive systems will be put through iterative testing activities, including playtesting and cognitive interviews. These iterative testing phases are especially important given unknown factors associated with these platforms for innovative technology-based items. NCES contracted the Educational Testing Service (ETS) to develop the platforms and associated items, and to carry out the necessary studies.

Volume I of this submittal describes the design, data collection, burden, cost, and schedules of the research activities for the aforementioned projects. Volume I Appendices provide recruitment and communication materials; and Volume II provides protocols and questions used in the research sessions.

### **Types of Research Methods**

The following sections describe the different types of research methodologies that will be used.

#### **Playtesting (used in early research phases)**

In playtesting, a process adapted from the game-design industry, a diverse set of students in small teams of two to four will work through and discuss activities, problems, and tasks with one another. An observer/facilitator will give overviews of the activities to students and provide guidance on what students should reflect on. Playtesting will take place early in the test development process using preliminary versions of the virtual systems. The purpose of playtesting is to gather student views on early versions of the interactive technology and begin to understand the range of ways that students use them. The primary goal here is evaluating and refining the platform and activities.

During playtesting, students will be encouraged to talk together about issues they confront, while observers note reactions to and potential problems with content or format. Observers will query students to draw them out, facilitate deeper reactions, or probe areas of possible confusion. Through playtesting, researchers will be able to identify construct-irrelevant features in tasks, such as inaccessible language, difficult interactions, or uninteresting or unfamiliar scenarios or activities that result in poor student engagement. Playtesting early in the research and development cycle allows for refinements to the system that can be tested in subsequent more intensive cognitive interviews. Playtesting will be only used in the Interactive Virtual Model project.

#### **Cognitive Interviews (used in the middle phases of the research project)**

In cognitive interviews (often referred to as a cognitive laboratory study or cog lab), an interviewer uses a structured protocol in a one-on-one interview drawing on methods from cognitive science. The objective is to explore how students are thinking and what cognitive processes they are using as they work through

tasks. The primary goal here is understanding how students think with the systems, and exploring what kinds of evidence of student cognition the systems can elicit.

Two methods will be combined: think-aloud interviewing and verbal probing techniques. With think-aloud interviewing, respondents are explicitly instructed to "think-aloud" (i.e., describe what they are thinking) as they work through questions or tasks. With verbal probing techniques, the interviewer asks probing questions, as necessary, to clarify points that are not evident from the "think-aloud" process, or to explore additional issues that have been identified a priori as being of particular interest. This combination of allowing students to verbalize their thought processes in an unconstrained way, supplemented by specific and targeted probes from the interviewer, has proven to be productive in previous NAEP pretesting<sup>1</sup> and will be the primary approach in the NAEP cognitive interviews described in this package.

Cognitive interview studies produce largely qualitative data in the form of verbalizations made by students during the think-aloud phase or in response to the interviewer probes. Some informal observations of behavior are also gathered, since typically a second observer is involved, in addition to the interviewer. Behavioral observations may include such things as nonverbal indicators of affect, suggesting emotional states such as frustration or engagement, and interactions with the task, such as ineffectual or repeated actions suggesting misunderstanding or usability issues. Cognitive Labs are used in both the Interactive Virtual Model and the Virtual Science Lab.

### **NAEP SAIL Research & Development: Technology Based**

Given that these SAIL projects involve technology-based platforms, all of the research activities will be conducted using technology (e.g., tablet, computer, or game-like control devices)<sup>2</sup>. Playtesting will use preliminary versions of the systems, while cognitive interviews will be conducted using interim versions.

## **3) Sampling and Recruitment Plans**

### **Playtesting**

Students will be recruited from districts that are located near the ETS campus, in Princeton, New Jersey, for scheduling efficiency and flexibility. ETS will recruit students, representing a range of demographic groups, using existing ETS contacts with individual parents/guardians, as well as administrators, teachers, and staff at local urban and suburban schools and afterschool programs for students. In some cases ETS will directly contact parents/guardians of students who have previously participated in ETS research and who are known to fit the targeted range of grade level, gender, race/ethnicity, socioeconomic background, and district type (urban, suburban, rural). In other cases, flyers, email, or letters will be used to contact parents/guardians and administrators or teachers/staff. School administrators/teachers may be asked to distribute paper flyers to students and parents. During these communications, the parent/guardian will be informed about the objectives, purpose, and participation requirements<sup>3</sup> of the data collection effort, as well as the activities that it entails. Confirmation emails and/or letters will be sent to participants. Only after ETS has obtained written consent from the parent/guardian will the student be allowed to participate in playtesting. See appendices A-I for representative recruitment, consent, confirmation, and thank you

<sup>1</sup> For example, NAEP Science Pretesting Activities (OMB #1850-0803 v.73, October 2012) and NAEP 2011 Cognitive Interview Studies of NAEP Cognitive Items (OMB #1850-0803 v.45, March 2011).

<sup>2</sup> For the ease of description, the term "computer" has been used in the recruitment materials.

<sup>3</sup> Screening questions will be used in the recruitment process to obtain the proper mix of students (appendices C and D).

materials.

A small number of students will participate in playtesting (see Table 1 for exact numbers for each project). A small sample is sufficient at the playtesting stage given that the key purpose is to identify usability errors and other construct-irrelevant issues.<sup>4</sup> The following will be recruited for playtesting:

- SAIL Virtual Science Lab: 10 students total from 8<sup>th</sup>/9<sup>th</sup> Grades.

### **Cognitive Laboratories**

To the extent possible, given the small sample size, students will be recruited by ETS staff for the SAIL cognitive laboratories/interviews (a.k.a. cog labs) from the following demographic populations:

- A mix of race/ethnicity (Black, Asian, White, Hispanic, etc.);
- A mix of socioeconomic background; and
- A mix of urban/suburban/rural.

Although the sample will include a mix of student characteristics, due to the small sample sizes the results will not explicitly measure differences by those characteristics. The recruitment process for the cognitive interviews will be the same as described above for playtesting. The materials in appendices A through I will also be used in the cog lab recruitment process. The numbers to be recruited for cog labs are:

- SAIL Interactive Virtual Models: Up to 40 students total from 8<sup>th</sup>/9<sup>th</sup> Grades, and
- SAIL Virtual Science Lab: Up to 15 students total from 8<sup>th</sup>/9<sup>th</sup> Grades.

Several researchers have confirmed the standard of five as the minimum number of participants per subgroup for analysis for the purposes of exploratory cognitive interviewing.<sup>5</sup> Thus, the numbers we have suggested should be sufficient for cognitive interviews given that the activities involve some complexity.

Table 1 summarizes the number of students to participate in the playtesting and cognitive interviews.

**Table 1. Sample Size: Cognitive Pretest Activities: Playtesting and Cognitive Interviews**

	<b>Virtual Models</b>	<b>Virtual Science Lab</b>	<b>Total</b>
Playtesting	0	10	10
Cognitive Interview	40	15	55
<b>Total</b>	<b>40</b>	<b>25</b>	<b>65</b>

## **4) Data Collection Process**

### **Playtesting**

Playtesting will take place at the ETS campus, in one of three dedicated research laboratories that are set up with recording equipment and working space for observers, facilitators, and one or more students (suitable for individual or small group sessions). Participants will first be welcomed and introduced to the facilitators/observers (assessment specialists, cognitive scientists, research assistants/associates, or task

<sup>4</sup> Nielson, J. (1994). Estimating the number of subjects needed for a think aloud test. In J. Human-computer Studies. 41, 385-397. Available at: <http://www.idemployee.id.tue.nl/g.w.m.rauterberg/lecturenotes/DG308%20DID/nielsen-1994.pdf>

<sup>5</sup> Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). The think-aloud method: A practical guide to modeling cognitive processes. San Diego, CA: Academic Press. Available at: <http://staff.science.uva.nl/~maarten/Think-aloud-method.pdf>

designers), and will be reassured that their participation is voluntary and that their answers may be used only for research purposes (see Section 6). Observers will then give an overview of the planned activities to students and provide guidance about what students should focus on. Observers will take notes on what students say and the sessions will be audio recorded. In addition, where feasible, screen-capture (e.g., *Camtasia* or *Morae Recorder*) will be used to record the actions occurring on the screen; note that this screen recording will not provide any identifiable data about the student. If log file capture (i.e., a digital record of all interactions with the system) is available, all student actions with the system will also be recorded in a data file; this will not provide any identifiable data, since students will be coded with an anonymous ID number. The SAIL Virtual Models project will also use digital video capture of the student interacting with the system, since their movements with the controller(s) will be an important part of the data. Parents and students will be informed about the video recordings prior to the sessions and their informed consent will be part of the criteria for participation.

For the most part, students will be allowed to explore and interact with the mocked-up task and activities by themselves with little intrusion on the part of the observer. However, at a few strategic points, observers may introduce questions meant to explore students' reactions to the task, areas of confusion, and ways of thinking about answers to the questions in the tasks and/or items. Examples of such questions are

- *Did you find the problem in this task interesting – why or why not?*
- *Are there any aspects of this that are confusing? Did you understand that part?*
- *How would you answer this question/How would you do this activity?* [Ask different group members if their approaches would differ]
- *How could this task/activity/system be improved? Could it be clearer? Could it be more interesting? Could it be easier to interact with?*

Prior to each playtesting session, ETS staff may identify some key focus areas for activity or for the system that students will be working with. If students do not provide sufficient comments on targeted parts, an observer may ask a group of students if they had any thoughts about the particular sections, using questions such as those described above, but focused on specific places or issues in the task or activities or system. See Volume II, Part B for the protocol used in the study.

### Analysis Plan

Feedback from a playtesting session is immediate and can be evaluated after the session. Notes from the observers in each session will be aggregated; one aggregate document will be produced for each task or set of items that are observed, with all observers contributing their observations to this common document. Since playtesting is a more informal process that generates relatively unstructured information, no formal quantitative analyses of these data will be performed, and the qualitative analyses will seek to pick out themes or individual observations that are important for the goals of developing the system or tasks going forward.

### Cognitive Laboratories

Cognitive laboratories will take place at the ETS campus, in one of the dedicated research laboratories described in the section on playtesting above. All sessions will be individual, and will be attended by two ETS staff (typically a facilitator/interviewer and an additional observer).

Participants will be welcomed, introduced to the interviewer and the observer, and told they are there to help develop and try out new kinds of systems for gathering information about students. They will be reassured that their participation is voluntary and that their answers may be used only for research purposes (see Section 6). Interviewers will explain the think-aloud process and conduct a practice session with a sample question.

On completion of the think-aloud component, the interviewer will proceed with follow-up questions (examples can be found in Volume II, Part C III). In this verbal probing component, the interviewer asks the student targeted questions about specific aspects of knowledge, skill, or ability that the task is attempting to measure, so that the interviewer can collect more information on the strategies and reasoning that the student employed as he or she worked through the task. The targeted questions will be generated for each task prior to testing. The interviewer is also encouraged to raise additional issues that became evident during the course of the interview. For example, if a student paused for a long time over a particular section, appeared to be frustrated at any point, or indicated sudden realization, the interviewer might probe these kinds of observations further, to find out what was going on. The interviews will be based on the protocol structures described in Volume II, Part C III.

As with the playtesting sessions, observers will take notes on what students say, and the student's think-aloud verbalizations will be captured using digital audio recording. Screen-capture software will be used to record the actions occurring on the screen. Additionally, for the Virtual Science Lab and the Virtual Models hybrid hands-on task, a webcam mounted on the tablet will record student interactions on the tablet surface. The combination of screen-capture and video of touch gestures is important in order to gather information about all of the actions a student may have made, including those that did not result in a change on screen (e.g., unsuccessfully attempting to apply an interactive gesture that was not recognized by the system or attempting to interact with a non-interactive element). For the Virtual Models project, a webcam mounted on a tablet will also employ visual object tracking technology in the hybrid hands-on task, but video recordings will capture the student interacting with the system in both tasks since their movements with the controller(s)/natural objects will be an important part of the data. Parents and students will be informed about the audio and video recordings prior to the sessions and their informed consent to being recorded will be part of the criteria for participation. If log file capture is available by this point in the projects, student actions with the systems will also be recorded in a data file (neither of which will produce identifiable data about the student). These recordings can be replayed or analyzed later, to see how a given student progressed through the task and what actions they took. Interviewers will also record their own notes separately, including behaviors (e.g., the participant appeared confused), whether extra time was needed, whether prior knowledge was evident, and so on.

### Analysis Plan

For the cognitive laboratory data collections, documentation will be grouped at the activity level. The types of data collected about each activity will include

- think-aloud verbal reports;
- behavioral data (e.g., actions observable from screen-capture or video of student);
- responses to generic questions prompting students to think out loud;



- responses to targeted questions specific to the activity;
- additional volunteered participant comments; and
- debriefing questions.

The general analysis approach will be to compile the different types of data to facilitate identification of patterns of responses for specific tasks or activities, such as patterns of frequency counts of verbal report codes and of responses to probes or debriefing questions, or types of actions observed from students at specific points in a given task. This overall approach will help to ensure that the data are analyzed in a way that is thorough, systematic, and that will enhance identification of problems with the systems or tasks and provide recommendations for addressing those problems.

## **5) Consultations Outside the Agency**

ETS will develop the platforms and associated items, perform recruitment and data collection activities, and carry out the necessary studies.

## **6) Assurance of Confidentiality**

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 consent will be obtained from parents or legal guardians of students who are under the age of 18. 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 consent forms, which include the participant name, will be separated from the participant interview files and secured for the duration of the study and will be destroyed after the final report is completed. Where sessions will be recorded<sup>6</sup>, the only identification included on the files will be the unique ID assigned to each participant by the interviewer. The recorded files will be secured for the duration of the study and will be destroyed when the research is complete.

## **7) Justification for Sensitive Questions**

Throughout the item and task development process, as well as the process of developing interview protocols, effort has been made to avoid asking for information that might be considered sensitive or offensive. Reviewers have attempted to identify and minimize potential bias in questions.

## **8) Estimate of Hourly Burden**

### Playtesting Burden

The estimated burden for recruitment assumes attrition throughout the process.<sup>7</sup> The anticipated total number of student participants for playtesting is 10. Around 15 teachers, school officials, and club and

<sup>6</sup> Recordings may be audio and/or video, as described in the specific interview sections.

<sup>7</sup> Assumptions for approximate attrition rates are 83 percent from initial contact (flyer from teacher) to screening form completion and 60 percent from submission of screening form to participation. (Note that slightly different attrition rates are estimated for the different pretesting phases.)

community center administrators will be contacted via email and phone. Initial email contact, response, and distribution of materials are estimated at 20 minutes or 0.33 hours. We anticipate distributing 150 flyers to parents and students via these 15 contacts. Time to review (for parents) is estimated at 5 minutes or 0.08 hours. Time to fill out the online screening form is estimated at 9 minutes or 0.15 hours. For those selected to participate and asked to fill out the consent form, the estimated time is 8 minutes or 0.13 hours. The follow-up email or letter to confirm participation for each session is estimated at 3 minutes or 0.05 hours. Playtesting sessions may last up to 90 minutes for all students. Table 2 details the estimated burden for playtesting.

**Table 2. Specific Burden for Playtesting studies<sup>8</sup>**

<b>Respondent</b>	<b>Hours per respondent</b>	<b>Number of respondents</b>	<b>Total hours</b>
<b>Student Recruitment via Teachers, Staff and club or community center administrators</b>			
Initial contact with staff: email, flyer distribution, and planning	0.33	15	5
<b>Parent or Legal Guardian, and Student</b>			
Flyer review	0.08	150	12
Completion of online screening form or phone screening	0.15	25*	4
Consent form completion and return	0.13	10**	2
Confirmation/acknowledgement to parent via email or letter	0.05	10**	1
<b>Recruitment Totals</b>		<b>165</b>	<b>24</b>
<b>Student</b>			
Virtual Science Lab	1.5	10	15
<b>Interview Totals</b>		<b>10</b>	<b>15</b>
<b>Total Burden</b>		<b>175</b>	<b>39</b>

\* Subset of initial contact group

\*\* Subset with completed screening forms

### Cognitive Interview Burden

The estimated burden for recruitment assumes attrition throughout the process.<sup>9</sup> The anticipated number of student participants for these cognitive interviews is 55 total. Around 48 school administrators and staff officials or club and community center administrators (and parents, if needed) will be contacted via email and phone. Initial email contact, response, and distribution of materials are estimated at 20 minutes or 0.33 hours. We anticipate distributing 647 flyers with consent forms via these 48 school and community group contacts to parents and students. Time to review is estimated at 5 minutes or 0.08 hours. Time to fill out the online screening form is estimated at 9 minutes or 0.15 hours. For those selected to participate and asked to fill out the consent form, the estimated time is 8 minutes or 0.13 hours. The follow-up email or letter to confirm participation for each session is estimated at 3 minutes or 0.05 hours. Individual cognitive interviews may last up to 90 minutes for all students. Table 3 details the estimated burden for the cognitive laboratories.

<sup>8</sup> The burden estimates in this table reflect the maximum burden for recruitment if students do not participate in multiple playtesting sessions.

<sup>9</sup> Assumptions for approximate attrition rates are 83 percent from initial contact (flyer from teacher) to screening form completion and 50 percent from submission of screening form to participation.

**Table 3. Estimate of Hourly Burden for Cognitive Interviews**

Respondent	Hours per respondent	Number of respondents	Total hours
Student Recruitment via School Administrators and Staff and club and or community center administrators			
Initial contact with staff: email, flyer distribution, and planning	0.33	48	16
Parent or Legal Guardian			
Flyer review	0.08	647	52
Completion of online screening form or phone screening	0.15	110*	17
Consent form completion and return	0.13	55**	8
Confirmation/acknowledgement to parent via email or letter	0.05	55**	3
<b>Recruitment Totals</b>		<b>695</b>	<b>96</b>
Student			
Interactive Virtual Models	1.5	40	60
Virtual Science Lab	1.5	15	23
<b>Interview Totals</b>		<b>55</b>	<b>83</b>
<b>Total Burden</b>		<b>750</b>	<b>179</b>

\* Subset of initial contact group

\*\* Subset with completed screening forms

#### Total for All Pretesting Activities

The combined totals for all of pretesting activities are listed in Table 4.

**Table 4. Combined Burden for SAIL Science Research Activities**

Pretest Activity	Number of respondents	Number of responses	Burden Hours
Playtesting	175	220	39
Cognitive Interviews	750	970	179
<b>Overall Totals</b>	<b>925</b>	<b>1,190</b>	<b>218</b>

## **9) Estimate of Costs for Recruiting and Paying Respondents**

To encourage participation and thank participants for their time and effort, a \$25 VISA gift card will be offered to each participating student, plus a \$25 VISA gift card to a parent or legal guardian bringing the student to and from the testing site.

## **10) Costs to Federal Government**

The estimated costs for the research activities in this submittal are described in Table 5.

**Table 5. Estimate of Costs**

Component	Provider	Estimated Cost
<b>SAIL Interactive Virtual Models</b> Design, prepare for, and carry out cognitive interviews (including recruitment, incentive costs, data collection, and summaries of findings)	ETS	\$294,104
<b>SAIL Virtual Science Lab</b> Design, prepare for, and carry out playtesting and cognitive interviews (including recruitment, incentive costs, data collection, and summaries of findings)	ETS	\$273,926
<b>Total</b>		<b>\$568,030</b>

## **11) Schedule**

Table 6 depicts the high-level schedule for the various activities. Each activity includes recruitment, data collection, analyses, and reports.

**Table 6. High-Level Schedule of Milestones**

<b>Activity</b>	<b>Dates</b>
Playtesting for Virtual Science Lab	February-August 2016
Cognitive labs for Virtual Science Lab	April- August 2016
Cognitive labs for Virtual Models	February -August 2016