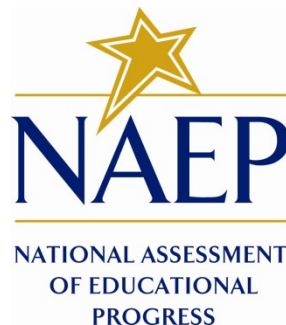


NATIONAL CENTER FOR EDUCATION STATISTICS
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Volume I
Supporting Statement

*NAEP Mathematics, Reading, and Science Hands-on Tasks
and National Indian Education Study (NIES) Survey
Pretesting*

OMB# 1850-0803 v.88
Play Testing, Cognitive Interviews, and Tryouts



November 12, 2013

Table of Contents

1)	Submittal-Related Information.....	1
2)	Background and Study Rationale.....	1
3)	Sampling and Recruitment Plans.....	4
4)	Data Collection Process.....	8
5)	Consultations Outside the Agency.....	13
6)	Assurance of Confidentiality.....	13
7)	Justification for Sensitive Questions.....	14
8)	Estimate of Hourly Burden.....	14
9)	Estimate of Costs for Recruiting and Paying Respondents.....	19
10)	Costs to Federal Government.....	19
11)	Schedule.....	20

1) Submittal-Related Information

This material is being submitted under the generic National Center for Education Statistics (NCES) clearance agreement (OMB #1850-0803). This generic clearance 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.

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 item development process, a portion of assessment items (cognitive and survey) are pretested on a small number of respondents before they are administered to a larger sample through pilot or operational tests. These pretest activities can include play testing and cognitive interviews, as well as tryouts of items, as defined later in this section. As paper-and-pencil administered NAEP assessments transition to technology-based assessments (TBA), new technology-enhanced items and scenario-based tasks (SBTs¹) will be developed featuring a range of possible designs. Pretesting is especially important given unknown factors associated with innovative technology-based items. NCES contracted the Educational Testing Service (ETS) to carry out the pretesting.

This submittal requests clearance for various pretesting activities related to the upcoming assessments. Specifically, these pretest activities are:

- Play testing for mathematics and reading/English language arts (ELA) selected items and tasks (to be administered in 2016) with students at grades 4 and 8 and science hybrid hands-on tasks (HOTs)² (to be administered in 2015) with students at grades 4, 8, and 12;
- Cognitive interviews for the mathematics and reading/ELA selected items and tasks with students at grades 4 and 8;
- Small-scale tryouts for the mathematics and reading/ELA selected items and tasks with students at grades 4 and 8;
- Cognitive interviews for the 2015 National Indian Education Study (NIES) survey questions with students, teachers, and school administrators at grades 4 and 8.

Included in the submittal are:

- Volume I — supporting statement that describes the design, data collection, burden, cost, and schedules of the pretesting activities for the aforementioned assessments;
- Volume I Appendices — recruitment and communication materials; and
- Volume II — protocols and questions used in the pretesting sessions.

Types of Pretesting

¹ SBTs are extended performance tasks, which embed multiple items into a scenario, providing context and motivation.

² Science hybrid hands-on tasks involve physical manipulatives along with tablet-delivered directions, questions, and response input.

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

Play Testing (used in pretesting the cognitive items)

In play testing, 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 scenario-based tasks and small sets of technology-enhanced items and tasks with one another. An observer/facilitator will give overviews of the tasks to students and provide guidance on what students should reflect on while looking at the tasks. Play testing will take place early in the test development process using preliminary versions of draft tasks. The purpose of play testing is to gather student views on early versions of interactive technology-based tasks and begin to understand how students are thinking about those tasks.

During play testing, students will be encouraged to talk together about items or tasks and 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 play testing, researchers will be able to identify construct-irrelevant features in tasks, such as inaccessible language or uninteresting or unfamiliar scenarios that result in poor student engagement. Play testing early in the development cycle also allows for task refinements that can be tested in subsequent and more intensive cognitive interviews.

Cognitive Interviews (used in pretesting the cognitive and NIES survey items)

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 reasoning processes they are using as they work through tasks. 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 pretesting³ and will be the primary approach in the NAEP cognitive interviews under 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.

In addition to think-aloud and verbal probing techniques, eye tracking methodology may be used during cognitive interviews for the cognitive items. Using this methodology, the student's gaze is tracked as he

³ 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).

or she works through a task, and the resulting eye movements can be interpreted to infer attentional and reasoning processes. In previous studies, eye tracking has been combined with a retrospective think-aloud method: students completed the task uninterrupted, and then watched the visible trace of their own gaze patterns overlaid onto the task as it was replayed to them in real time. This gaze trace, along with their own mouse clicks and responses, helped students reconstruct their thinking as it had previously occurred, allowing them to “think aloud” retrospectively. Eye tracking provides a unique opportunity to gather data about how students process tasks; it does not require explicit probing or for students to articulate their thought processes.

Small-Scale Tryouts (used in pretesting the cognitive items)

During small-scale tryouts, students work uninterrupted through a selected set of draft programmed items or tasks. The strength of using a tryout methodology on a small scale is that it allows data to be gathered about student responses and actions during normal, uninterrupted item or task performance. This approach provides a small-scale snapshot of the ranges of responses and actions that items and tasks are meant to elicit, but which can be gathered much earlier in the assessment development process and with fewer resource implications than formal piloting. Previous experience, for example with the NAEP Technology Engineering Literacy Assessment⁴, shows that tryout-based insights are very informative, especially for the refinement of scoring rubrics (e.g., for examining, characterizing, and grouping the types of constructed responses that students provide and allocating appropriate scoring levels accordingly) and for finalizing or revising decisions about student actions to be captured.

NAEP Technology-based Assessments in Mathematics, Reading/ELA, and Science

Given that the assessments will be technology-based, all of the pretesting activities will be conducted using technology (e.g., a tablet or laptop)⁵. Play testing will use preliminary versions of draft tasks, while cognitive interviews will be conducted using draft programmed tasks, and small-scale tryouts will be conducted on the final versions.

In 2016, new technology-enhanced items and SBTs for mathematics and reading will be piloted for use in operational NAEP assessments. New assessment content in mathematics and reading will employ instruments designed to deepen and expand measurement of framework content and explore innovative ways of measuring subject knowledge and skills. All three types of pretesting will be conducted with information learned from one stage informing item development and revisions that are then tested in a subsequent stage.

Pretesting for science hybrid HOTs, which will be pilot tested in 2015, is intended to explore using technology to enhance and improve the existing paper-based HOTs. For example, among the enhancements that will be pretested are using technology to replace lengthy paper-based directions with video-based directions, examples, and demonstrations; task text “chunking” (delivering pieces of text for students to digest in parts versus all at once) to reduce reading burden; and midstream “correction” achieved by giving students accurate data or feedback once their responses to items have been submitted

⁴ Technology and Engineering Literacy Pre-Assessment Studies: Tryout and Usability Studies (OMB #1850-0803 v.66, February 2012).

⁵ For the ease of description, the term “computer” has been used in the recruitment materials.

(thus allowing students to go further in tasks).

Survey Questions – NIES

In addition to the main NAEP core and subject-specific survey questionnaires, the NIES survey is administered to students in grades 4 and 8 who are identified as American Indian/Alaska Native (AI/AN), their reading and mathematics teachers, and their school administrators. The student, teacher, and school NIES questionnaires provide culturally-specific contextual information for in-depth reporting on the academic achievement and educational experiences of AI/AN students in grades 4 and 8. The survey is focused on the integration of native language and culture into school and classroom activities. The NIES survey was last administered in 2011 and will be administered in 2015.

Periodically, NCES will add, revise, or delete questions from existing survey questionnaires. These modifications aim to improve questionnaire quality, replace or drop outdated questions, and collect data on new contextual factors that are expected to be associated with academic achievement. Questionnaires for the 2015 NIES assessment have undergone a systematic review process that has resulted in the addition and revision of several questionnaire items. These items have been reviewed by expert panels and NCES. Prior to the 2015 NIES assessment, these new and revised items, as well as some trend items, will undergo cognitive interview testing. In addition, a subset of existing items will be included to enable comparison between the existing items and the proposed new or revised items. The 2015 NIES assessment will be administered as a paper-and-pencil assessment. Therefore, the cognitive interviews for NIES will be conducted using paper-based instruments.

3) Sampling and Recruitment Plans

The sampling and recruitment plans, which differ by the type of testing, are described below.

Play Testing (Mathematics, Reading/ELA, and Science Hybrid HOTs)

Educational Testing Service (ETS) will conduct the play testing. 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 teachers and staff at local urban and suburban schools and afterschool programs for students. E-mail or letters will be used to contact these teachers/staff; who will then distribute paper flyers and consent forms to students and parents. During this communication, the parent/guardian will be informed about the objectives, purpose, and participation requirements of the data collection effort, as well as the activities that it entails. Confirmation e-mails 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 the play testing session. See appendices A through I for representative recruitment, confirmation, and thank you materials.

Five students per grade will participate and provide feedback for each play testing task; five students should be sufficient at the play testing stage given that the key purpose is to identify usability errors and other construct-irrelevant issues.⁶ For mathematics and reading/ELA, each play testing session will include at least one SBT and optionally some selected technology-enhanced discrete items (i.e., items that

⁶ 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>

are not associated with an SBT). Given the total number of tasks being developed, each subject will have up to 3 play testing sessions for each grade (4 and 8), for a total of up to 6 math and 6 reading/ELA sessions. Based on prior experience with similar studies, it is anticipated that some of the same students will return to participate in multiple sessions. Therefore, play testing is expected to involve a minimum of 20 and a maximum of 60 students across the two grades and subjects.

For science hybrid HOTs, each play testing session will include at least one task. Given the total number of tasks being developed, two play testing sessions at each grade (4, 8, and 12) will be planned for a total of 6 sessions. Based on prior experience with similar studies, it is anticipated that some of the same students will return to participate in multiple sessions. Therefore, play testing is expected to involve a minimum of 15 and a maximum of 30 students across the three grades for hybrid HOTs.

Cognitive Laboratories (Mathematics and Reading/ELA)

ETS will conduct the cognitive interviews for mathematics and reading/ELA. Students will be recruited by ETS staff 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, the results will not explicitly measure differences by those characteristics. Students will be recruited from districts that are located near the ETS Princeton, New Jersey campus for scheduling efficiency and flexibility. As with play testing, ETS will recruit students using existing ETS contacts with administrators and staff at local schools and afterschool programs for students. If needed, ETS may also reach out directly, via e-mail, letter, or phone, to parents. E-mails, letters, or phone calls will be used to contact administrators and staff at local schools and afterschool programs. Paper flyers and consent forms for students and parents will be distributed through these school administrators and staff contacts. The parent/guardian will be informed about the objectives, purpose, and participation requirements of the data collection effort, as well as the activities that it entails. Confirmation e-mails and/or letters will be sent to participants. Only after ETS has obtained written consent from the parent/guardian will a student be allowed to participate in the cognitive interview session. See appendices J-V for representative recruitment, consent, confirmation, and thank you materials.

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.⁷ A sample size of 5 to 15 individuals has become the standard for NAEP cognitive interviews.⁸ Based on this research and prior experience, seven to ten students per task, per grade, and subject should be sufficient for cognitive interviews given that the tasks involve some complexity. Based on the number of tasks that can be

⁷ 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://akmc.biz/ShareSpace/ResMeth-IS-Spring2012/Zhora_el_Gauche/Reading%20Materials/Someren_et_al-The_Think_Aloud_Method.pdf

⁸ For example, NAEP Science Pretesting Activities (OMB #1850-0803 v.73, October 2012) and Cognitive Interview Study of Background Questions for Students, Teachers, and School Administrators (OMB #1850-0803 v.57, September 2011).

completed per session and the number of tasks to go through the cognitive interview process, cognitive interviewing is expected to involve a maximum total of 120 students across grades 4 and 8 for mathematics and reading/ELA.

Small-scale Tryouts for Mathematics and Reading/ELA

EurekaFacts will perform the tryouts, recruiting from the greater Washington, DC/Baltimore metropolitan area, ensuring the results are representative of various populations. Students will be sampled to obtain the following:

- 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, the results will not explicitly measure differences by those characteristics.

While EurekaFacts will use various outreach methods to recruit students to participate, the bulk of the recruitment will be conducted by telephone and based on their acquisition of targeted mailing lists containing residential address and land line telephone listings. EurekaFacts will also use a participant recruitment strategy that integrates multiple outreach/contact methods and resources such as newspaper/Internet ads, outreach to community organizations (e.g., Boys and Girls Clubs, Parent-Teacher Associations), and mass media recruiting (such as postings on the EurekaFacts website).

Interested students will be screened to ensure that they meet the criteria for participation in the tryout (e.g., their parents/guardians have given consent and they are from the targeted demographic groups outlined above). When recruiting participants, EurekaFacts staff will first speak to the parent/guardian of the interested minor before starting the screening process. During this communication, the parent/guardian will be informed about the objectives, purpose, and participation requirements of the data collection effort as well as the activities that it entails. After confirmation that participants are qualified, willing, and available to participate in the research project, they will receive a confirmation e-mail/letter and phone call. Informed parental consent will be obtained for all respondents who are interested in participating in the data collection efforts. (See appendices W-AI for representative tryout recruitment, consent, confirmation, and thank you materials.)

EurekaFacts will recruit 25 students for each scenario-based task. In addition to the SBT, students may take selected technology-enhanced discrete items. Up to 200 students will be recruited for small-scale tryouts across grades 4 and 8 for mathematics and reading/ELA. Students will participate in tryouts either individually or in groups. Table 1 summarizes the number of students for the play testing, cognitive interviews, and tryout components of the cognitive pretesting activities.

Table 1. Sample Size: Cognitive Pretest Activities: Play Testing, Cognitive Interviews, Tryouts⁹

	Grade 4	Grade 8	Grade 12	Total
Play Testing	40	40	10	90
Cognitive Interview	60	60	NA	120
Tryouts	100	100	NA	200
Total	200	200	10	410

NIES Cognitive Interviews

Drawing on existing research and practices, the following sections describe the sampling processes that will be utilized for the cog labs for the grade 4 and grade 8 NIES survey questionnaires for students, teachers, and school administrators. The NIES survey is more complex than general NAEP survey questionnaires due to the high geographical and cultural diversity of the AI/AN population. The goal of the cognitive interviews is to identify potential problems of newly developed or revised questions as they apply to the specific population. While the cognitive interviews for cognitive items are one stage in a series of pretesting activities prior to an operational assessment, the cognitive interviews for the NIES survey is the only pretesting activity. Therefore, the cognitive interviews will receive more prominence in helping to ensure the validity and adequacy of the proposed items. Therefore, to cover the diverse population and allow for the possibility of reacting to intermediate feedback after a first set of cognitive interviews, particularly for students, while maintaining the necessary sample sizes, we will conduct extended cognitive interviews with larger sample sizes than cognitive interviews for cognitive items.

To ensure a diverse sample, students will be sampled to obtain the following criteria:

- Representative mix of regions in states reflecting the AI/AN population (previously sampled regions for 2008 NIES cognitive interviews were: North Carolina, Wisconsin, Missouri, New Mexico, Alaska, and Montana)
- Mix of students attending different types of schools (e.g., public versus run by the Bureau of Indian Education (BIE); urban schools versus schools on tribal land)
- Mix of gender
- Mix of relevant AI/AN ethnicities
- Mix of socioeconomic backgrounds as applicable to AI/AN population

To ensure a diverse sample, teachers and school administrators will be sampled with the following criteria:

- School populations include fourth- and/or eighth-grade students
- Mix of gender
- Mix of relevant AI/AN ethnicities
- Mix of school sizes
- Mix of school socioeconomic demographics as applicable to AI/AN population
- Mix of different types of schools (e.g., public versus BIE run; urban schools versus schools on tribal land)

⁹ This table represents the expected distribution across grades. Depending on the nature of the items and tasks and the specific recruitment challenges, the actual distribution may vary slightly. For burden purposes, the maximum number of students by pretesting activity will not exceed the total shown in the table.

- Mix of experienced and less experienced teachers and school administrators
- Mix of mathematics and reading teachers
- Representative mix of regions in states reflecting the AI/AN population (previously sampled regions for 2008 NIES cognitive interviews were: North Carolina, Wisconsin, Missouri, New Mexico, Alaska, and Montana)

Although the sample will include a mix of these characteristics, the results will not explicitly measure differences by those characteristics.

Kauffman & Associates Incorporated (KAI) will perform the cognitive interviews. KAI will contact tribal leaders, school administrators, teachers, and parents or legal guardians of AI/AN students via letter, email, and/or telephone to recruit participants for the survey questionnaire interviews. Before conducting research with a sovereign nation, KAI reaches out to tribal leadership via letter, email, and/or telephone to inform them of the interest to conduct research with the tribe's population and on the tribe's land. Researchers working in Indian Country are accountable to the tribes with whom they work and respect and honor that process by reaching out to tribal leadership before seeking to conduct research.

Interested participants will be screened using a screener script to ensure the mix of characteristics described above. For selected participants, KAI will confirm the interview date, time, and location. School administrators, teachers, and parents or legal guardians of participating students will complete consent forms at the time of the interview. See appendices AJ-AY for representative introduction, recruitment, confirmation, consent, and thank you materials. Table 2 summarizes the number of students, teachers, and school administrators for the NIES cognitive interviews, based on the number of items to be tested.

Table 2. Sample Size: NIES Survey Questionnaire Items

	Grade 4	Grade 8	Total
Students	40	40	80
Teachers	20	20	40
School Administrators	10	10	20
Total	70	70	140

4) Data Collection Process

Play Testing (Mathematics, Reading, and Science Hybrid HOTS)

Play testing will take place in a range of locations so that staff can maximize opportunities to work with students. Depending on scheduling and participants, some could take place at ETS, some in schools (after school), and some at organizations from which students will be drawn (e.g., at Boys and Girls Clubs).

Participants will first be welcomed and introduced to the facilitators/observers (assessment specialists, cognitive scientists, or task designers), and will be reassured 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]. Observers will then give an overview of the tasks and/or items to students and provide guidance about what students should focus on while looking at the tasks. Observers will take notes on what students say and the sessions will be audio recorded.

For the most part, students will be allowed to explore and interact with the mocked-up task and item versions 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 questions or words that seem confusing here? Did you understand that part?*
- *How would you answer this question? [Ask different group members if their approaches would differ].*
- *How could this task be improved? Could it be clearer? Could it be more interesting?*

Prior to each play testing session, ETS staff may identify some key focus areas for each task. 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. See Volume II, Part B for the protocol used in the study.

Feedback from a play testing 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 play testing is a more informal process that generates relatively unstructured information, no formal analyses of these data will be performed.

Cognitive Laboratories (Mathematics and Reading/ELA)

Cognitive interviews will take place at a range of suitable venues. In some instances, students may be invited to the ETS campus and in other cases ETS research staff will travel to schools or after-school venues to interview students. If conducted at a school, the interviews may be conducted during school hours or after school, based on the preference of the school administrators. In all cases, an appropriate environment such as a quiet room will be used to conduct the interviews.

Participants will first be welcomed, introduced to the interviewer and the observer (if an in-room observer is present), and told they are there to help answer questions about how people answer mathematics, reading, or (if extended writing is involved) English Language Arts tasks. Students will be reassured 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]. Interviewers will explain the think-aloud process and conduct a practice session with a sample question.

The think-aloud component of the cognitive interviews will use either 1) a concurrent think-aloud method in which the student verbalizes his or her thoughts while working through the task, or 2) a retrospective think-aloud method during which students work through the task silently and then discuss their thoughts

about the task content while working through it again. The second approach may, technology allowing, also utilize eye tracking to help cue students as to what they were focusing on when they did the task the first time.

The methods also include a verbal probing component conducted after completion of the think-aloud portion for a given task component. The verbal probes include a combination of pre-planned task-specific questions, identified before the session as important, and ad hoc questions that the interviewer identifies as important from observations during the interview, such as clarifications or expansions on points raised by the student. To minimize the burden on the student, efforts are made to limit the number of verbal probes that can be used in any one session or in relation to any one task. The protocols will contain largely generic prompts to be applied flexibly by the interviewer to facilitate and encourage students in verbalizing their thoughts. For example: “What’s going on in your head right now?” and “I see you’re looking at the task [or screen/figure/chart/text]. What are you thinking?” The interviews will be based on the protocol structures described in Volume II, Part C.

As described in Section 2, eye-tracking may also be used in conjunction with the cognitive interviews. Eye-trackers use an infrared video image of the eyes to calculate gaze location in real-time, so that it is possible to see where on the screen the student is looking at any given moment. First, students work through a task without interruption. During this phase their eye movements are unobtrusively recorded and all events on the screen are captured in real time. Next, students are asked to go back over the task a second time and attempt to introspect on, and describe out loud, what they were thinking as they went along. To help students recall their thinking as it occurred, a screen capture video of the complete task is replayed, including mouse movements and item responses, with a moving cursor showing the student’s gaze-patterns overlaid onto these screen images. Therefore, students can see exactly where and how they were looking while they were doing the task the first time around. Seeing their own eye movements acts as a prompt, helping test-takers reconstruct their thinking at each point in the task.

On completion of a task, the interviewer will proceed with follow-up questions. 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.

Interactions and responses will be recorded via video screen-capture software (e.g., *Camtasia*), as well as possibly via digital video of the student. These recordings can be replayed for later analysis, to see how a given student progressed through the task. The combination of the screen-capture and the video is important to determine all of the actions a student may have made 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). Digital audio recording will capture

students' verbal responses to the think-aloud interview. Interviewers will also record their own notes separately, including behaviors (e.g., the participant appeared confused) and whether extra time was needed during a particular part of the task.

Analysis Plans

For the cognitive interview data collections, documentation will be grouped at the task level. Task items will be analyzed across participants.

The types of data collected about task items and components will include

- think-aloud verbal reports;
- behavioral data (e.g., errors in reading items or tasks; actions observable from screen-capture; interactive gestures captured via webcam);
- responses to generic questions prompting students to think out loud;
- responses to targeted questions specific to the item or task;
- 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 items or tasks, 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 items or tasks and provide recommendations for addressing those problems.

Small-Scale Tryouts (Mathematics and Reading/ELA)

These studies will be conducted by EurekaFacts, which will recruit participants, conduct and observe the sessions, record interactions as appropriate, and report results to ETS. EurekaFacts will conduct tryouts at their Rockville, Maryland site. When working with students individually, EurekaFacts will screen capture the student actions using a software program such as Morae[®] (by TechSmith). Morae Recorder's core strength is its facility for capturing a student's interactive behaviors as they happen, while one or more observers simultaneously record text comments that are time-locked to the student actions and to the video recording. Adding Morae Observer software allows observers to be located in a remote location. This is both a convenience for observers and a potential means of reducing student stress or distraction, which can detract from data quality.

As with the cognitive interviews, if possible (i.e., if compatibility issues allow) the actions on the tablet screen will be recorded using screen-capture software (e.g., Morae), and the student gestures on the tablet surface may also be captured by a webcam attached and focused on the surface. Both video feeds will be directed to Morae to allow remote viewing and commenting, and later playback through this software, also assuming compatibility. In contrast to the cognitive interviews, in the tryouts there will be no think-aloud or verbal probing component, although students will be asked generic debriefing questions to get their overall impressions of tasks. Again, the goal of tryouts is to gather authentic, uncontaminated task performance and action data. Therefore, students will work through tasks and selected items at their own pace and without interruption. The protocol is described in Volume II, Part D.

Analysis Plan

Student responses to items will be compiled into spreadsheets to allow quantitative and descriptive analyses of the performance data. For the behavioral data, the videos will be used for qualitative analysis to characterize the range of behaviors observed for tryouts that are conducted with one student at a time. Once the coding is established, a basic quantitative analysis will provide frequency counts and, where relevant, order information, for different behaviors or actions observed from each student. These will also be compiled into spreadsheets, and the performance data and behavioral data for each student will be combined in the same document.

Cognitive Interview Process for NIES Survey Questionnaires

KAI will conduct the cognitive interviews, with oversight from NCES and ETS. ETS will train KAI staff members on how to administer the interview probes and document the data in specified formats. KAI will ensure that qualified interviewers, trained on the cognitive interviewing techniques of the protocols, conduct the interviews. The interviews will be based on the protocol structures described in Volume II (Part E).

Participants will first be welcomed, introduced to the interviewer and the observer (if an in-room observer is present), and told they are there to help answer questions about how people respond to survey items. Participants will be reassured that their participation is voluntary and that their responses may be used only for research purposes and will 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]. Interviewers will explain the think-aloud process, conduct a practice question, and then participants will answer questions verbally.

Digital audio recording will capture students' verbal responses to the think-aloud interview. Interviewers may also record notes, including behaviors (e.g., the participant appeared confused) and if extra time was needed during a particular part of the task.

Student interviews will take place at AI/AN and BIE schools and/or community centers. Teachers and school administrators who agree to participate will be interviewed at their school locations. While cognitive interviews with students should be conducted as in-person interviews, half of the cognitive interviews with teachers and school administrators may be conducted via telephone if sampling requirements do not allow for testing participants in person, or if complications with transportation and/or reaching certain populations can be avoided. At least half of the in-person interviews should be completed prior to beginning the telephone interviews in order to create a better forum for identifying and correcting any major issues early on. For example, if, of the 20 school administrators that will be interviewed, 10 interviews will be conducted over the phone, at least five of those interviews will be conducted in-person before beginning any telephone interviews.

After the participant reads each question and answers it while thinking aloud, the interviewer will ask both item-specific probes, as well as possibly asking some generic probes. The protocol which contains the welcome script, think-aloud instructions, hints for the interviewers, the specific survey items included, and the generic and item-specific probes are contained in Volume II, Part E.

Analysis Plans

For the cognitive interview data collections, the key unit of analysis is the item. Items will be analyzed across participants.

The types of data collected about the survey questions will include

- think-aloud verbal reports;
- behavioral data (e.g., errors in reading items);
- responses to generic questions prompting students to think out loud;
- responses to targeted questions specific to the item;
- additional volunteered participant comments; and
- debriefing questions.

The general analysis approach will be to compile the different types of data in spreadsheets and other formats to facilitate identification of patterns of responses for specific items, for example, patterns of counts of verbal report codes and of responses to probes or debriefing questions. This approach will help ensure that the data are analyzed in a way that is thorough, systematic, and that will enhance identification of problems with items and provide recommendations for addressing those problems.

5) Consultations Outside the Agency

Educational Testing Service (ETS) serves as the Item Development contractor. As such, ETS will be responsible for the management of all activities described in this package.

EurekaFacts, located in Rockville, Maryland, is a small, established for-profit research and consulting firm, offering facilities, tools, and staff to collect and analyze both qualitative and quantitative data. EurekaFacts is working as a subcontractor for ETS to conduct the small-scale tryouts.

Kauffman and Associates, Inc (KAI) is an American Indian, woman-owned organization that has worked on issues of education, health, and well being of Native communities for over 20 years. KAI has worked closely with hundreds of tribal communities and thousands of tribal members to address a range of concerns. This company is known for its culturally appropriate sensitivity and knowledgeable work in support of Native communities, and is highly respected by tribal leaders for the services they provide.

6) Assurance of Confidentiality

Participants are 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 participants who are over the age of 18 and 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.

The interviews will be recorded¹⁰. 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 after the final report is submitted.

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

Play Testing Burden – Mathematics, Reading/ELA and Science HOTS

The estimated burden for recruitment assumes attrition throughout the process.¹¹ The anticipated number of student participants for play testing is 35-90 (given that some students may participate in multiple sessions). Teachers and school officials will be contacted via e-mail and phone. Initial e-mail contact, response, and distribution of materials are estimated at 20 minutes or 0.33 hours. We anticipate distributing 240 flyers with consent forms via these contacts to parents and students. Time to review flyers and consent forms is estimated at 5 minutes or 0.08 hours. For those choosing to fill out the consent form, the estimated time is 8 minutes or 0.13 hours. The follow-up e-mail or letter to confirm participation for each session is estimated at 3 minutes or 0.05 hours. Play testing sessions are expected to last 60 minutes for all students. Table 3 details the estimated burden for play testing.

¹⁰ Recordings may be audio and/or video, as described in the specific interview sections.

¹¹ Assumptions for approximate attrition rates are 50 percent from initial contact (flyer from teacher) to consent form completion and 25 percent from submission of consent form to participation.

Table 3. Specific Burden for Play Testing for Mathematics, Reading/ELA and Science HOTS¹²

Respondent	Hours per respondent	Number of respondents	Total hours
Student Recruitment via Teachers and Staff			
Initial contact with staff: e-mail, flyer distribution, and planning	0.33	12	4
Sub-Total		12	4
Parent or Legal Guardian, and Student (18 or older)			
Flyer and consent form review	0.08	240	19
Consent form completion and return	0.13	120*	16
Confirmation to parent via email or letter	0.05	90*	5
Sub-Total		240	40
Recruitment Totals		252	44
Student			
Grade 4	1	40	40
Grade 8	1	40	40
Grade 12	1	10	10
Interview Totals		90	90
Total Burden		342	134

* Subset of initial contact group (**total number of responses = 552**)

Cognitive Interview Burden – Mathematics and Reading/ELA

The estimated burden for recruitment assumes attrition throughout the process.¹³ The anticipated number of student participants for these cognitive interviews is 120 total. School administrators and staff officials (and parents, if needed) will be contacted via e-mail and phone. Initial e-mail contact, response, and distribution of materials are estimated at 20 minutes or 0.33 hours. We anticipate distributing 320 flyers with consent forms via school contacts to parents and students. Time to review flyers and consent forms is estimated at 5 minutes or 0.08 hours. For those choosing to fill out the consent form, the estimated time is 8 minutes or 0.13 hours. The follow-up e-mail or letter to confirm participation for each session is estimated at 3 minutes or 0.05 hours. Individual cognitive interviews are expected to last 60 and 90 minutes for grade 4 and grade 8 students, respectively. Table 4 details the estimated burden for the mathematics and reading cognitive laboratories.

¹² The burden estimates in this table reflect the maximum burden for recruitment if students do not participate in multiple play testing sessions.

¹³ Assumptions for approximate attrition rates are 50 percent from initial contact (flyer from teacher) to consent form completion and 25 percent from submission of consent form to participation.

Table 4. Estimate of Hourly Burden for Cognitive Interviews for Mathematics and Reading/ELA

Respondent	Hours per respondent	Number of respondents	Total hours
Student Recruitment via School Administrators and Staff and Parents			
Initial contact with staff: e-mail, flyer distribution, and planning	0.33	300	99
Sub-Total		300	99
Parent or Legal Guardian			
Flyer and consent form review	0.08	320	26
Consent form completion and return	0.13	160*	21
Confirmation to parent via email or letter	0.05	120*	6
Sub-Total		320	53
Recruitment Totals		620	152
Student			
Grade 4	1	60	60
Grade 8	1.5	60	90
Interview Totals		120	150
Total Burden		740	302

* Subset of initial contact group (**total number of responses = 1,020**)

Small-Scale Tryout Burden – Mathematics and Reading/ELA

The estimated burden for recruitment assumes attrition throughout the process.¹⁴ The anticipated number of student participants for small-scale tryouts is 200. Based on the proposed outreach and recruitment methods, we estimate initial respondent burden, regardless of the mode of initial interaction (e.g., a telephone recruiting call, receipt of a request to participate by postal mail, or receipt of an e-mailed message regarding the study), at 3 minutes or 0.05 hours. The follow-up phone calls to conduct participant screening and schedule the interviews are estimated at 9 minutes or 0.15 hours per family. The follow-up phone call and letter to confirm participation is estimated at 3 minutes or 0.05 hours. Tryouts are expected to last 60 minutes for each student. Table 5 details the estimated burden for the mathematics and reading small-scale tryouts.

¹⁴ Assumptions for approximate attrition rates for direct parent recruitment of students are 80 percent from initial contact to follow-up, 20 percent from follow up to confirmation and 10 percent from confirmation to participation.

Table 5. Estimate of Hourly Burden for Small-Scale Tryouts for Mathematics and Reading/ELA

Respondent	Hours per respondent	Number of respondents	Total hours
Parent and Student Recruitment			
Initial contact	0.05	1,390	70
Follow-up via phone, including consent form completion and return	0.15	556*	83
Confirmations	0.05	222**	11
Recruitment Totals		1,390	164
Student			
Grade 4	1	100**	100
Grade 8	1	100**	100
Interview Totals			200
Total Burden		1,390	364

* This includes both parents and students from 278 households

** Subset of initial contact group (**total number of responses = 2,368**)

Cognitive Interview Burden – Survey Questions

The estimated burden for recruitment assumes attrition throughout the process.¹⁵ Initial contact and response is estimated at 3 minutes or 0.05 hours. The follow-up phone call to screen participants and/or answer any questions the participants (or their parents or legal guardians) have is estimated at 9 minutes or 0.15 hours per participant. The follow-up to confirm participation is estimated at 3 minutes or 0.05 hours. All interviews will be scheduled for no more than 60 minutes. Table 6 details the estimated burden for the survey questionnaire cognitive interviews.

¹⁵ Assumptions for approximate attrition rates for direct participant recruitment are 50 percent from initial contact to follow-up, 30-35 percent from follow up to confirmation and 15-20 percent from confirmation to participation.

Table 6. Estimate of Hourly Burden – Recruitment and Participation for Survey Questionnaire Cognitive Interviews¹⁶

Respondent	Hours per respondent	Number of respondents	Total hours
Tribal Contact			
Initial contact	0.05	90	5
Follow-up via phone	0.15	45*	7
Sub-Total		90	12
Parent or Legal Guardian for Student Recruitment			
Initial contact	0.05	300	15
Follow-up via phone	0.15	150*	23
Consent & Confirmation	0.05	100*	5
Sub-Total		300	43
Teacher and School Administrator Recruitment			
Initial contact	0.05	200	10
Follow-up via phone or e-mail	0.15	100*	15
Consent & Confirmation	0.05	70*	4
Sub-Total		200	29
Participation (Interviews)			
Grade 4 Students	1	40	40
Grade 8 Students	1	40	40
Teachers	1	40*	40
School Administrators	1	20*	20
Sub-Total		80*	140
Total Burden		670	224

* Subset of initial contact group (total number of responses = 1,195)

Total for All Pretesting Activities

The combined totals for all of pretesting activities are listed in Table 7.

Table 7. Combined Burden for Pretesting Activities

	Number of respondents	Number of responses	Burden Hours
Cognitive items and tasks			
Total Play Testing Burden	342	552	134
Total Cognitive Interview Burden	740	1,020	302
Total Tryout Burden	1,390	2,368	364
NIES Survey items			
Total Cognitive Interview Burden	670	1,195	224
Overall Totals	3,142	5,131	1,024

¹⁶ Participants will only be contacted after receiving tribal approval.

9) Estimate of Costs for Recruiting and Paying Respondents

For all student pretesting activities held outside of school hours, a \$25 Visa gift card will be given to each student, and, if transportation is provided, a parent or legal guardian of each student will receive a gift card of \$25 to thank him or her for the time involved and to help offset the travel/transportation costs.

If the mathematics and reading cognitive interviews take place at schools during school hours, the \$25 gift cards will be given to the school administrators.

For the NIES Survey teachers and school administrator cognitive interviews, participants will be given a \$40 gift card for participating in the study.

10) Costs to Federal Government

The estimated costs for the pretesting activities in this submittal are described in Table 8.

Table 8. Estimate of Costs

Activity	Provider	Estimated Cost
Cognitive Item Play Testing Design, prepare for, and conduct play testing sessions (including recruitment, incentive costs, data collection, and summary of findings).	ETS	\$ 84,060
Cognitive Item Cognitive Interviews Design, prepare for, and conduct cognitive interviews (including recruitment, incentive costs, data collection, analysis, and reporting).	ETS	\$ 202,040
Cognitive Item Small-scale Tryouts Design, prepare for, and conduct scoring and analysis of tryouts. Prepare for and conduct tryouts (including recruitment, incentive costs, data collection, reporting).	ETS EurekaFacts	\$ 104,600 \$104,066
Survey Questionnaire Cognitive Interviews Design, preparation, and analysis for survey questionnaire cognitive interviews; cognitive interview training for KAI staff. Preparation and conduct of survey questionnaire cognitive interviews (including recruitment, incentive costs, data collection, analysis, and reporting) for NIES pretest activities.	ETS KAI	\$ 35,000 \$157,996
Total		\$687,762

11) Schedule

Table 9 depicts the high-level schedule for the various activities. Each activity includes recruitment, data collection, analyses, and reports. In addition, the commencement of activities is contingent upon OMB approval.

Table 9. High-Level Schedule of Milestones

Activity	Dates
Play testing for mathematics and reading	December 2013-December 2014
Play testing for science hybrid HOTs	December 2013-March 2014
Cognitive interviews for mathematics and reading	April 2014-December 2014
Small-scale tryouts for mathematics and reading	September 2014-March 2015
Cognitive interviews for NIES survey	December 2013-March 2014