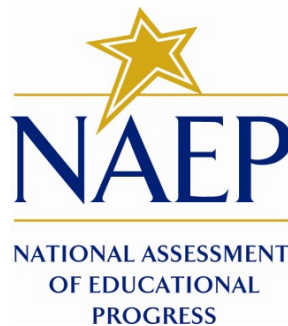# NATIONAL CENTER FOR EDUCATION STATISTICS
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Volume I
Supporting Statement

NAEP Mathematics and Reading Pretesting 2015

OMB# 1850-0803 v.140

August 4, 2015
Amendment to a previously approved package (1850-0803 v.88)

Table of Contents

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.

This package is an amendment to a previously approved package for the NAEP Mathematics, Reading, and Science Hands-on Tasks and National Indian Education Study (NIES) Survey Pretesting (1850-0803 v.88). The science hands-on tasks and NIES survey pretesting have been concluded. However, because the pilot test of the reading and math scenario-based tasks (SBTs[1]) has been rescheduled from 2016 to 2017, the pretesting window for these tasks has been expanded. Furthermore, the need for additional information to be gained from the tryouts requires an expansion of the sample size and inclusion of questions related to student experience with the subject area. This submission is a revision of the originally approved package (1850-0803 v.88) in that it only includes the components that have yet to be completed (reading cognitive interviews and reading and mathematics tryouts) with some changes to (1) the expected number of respondents and (2) the interview protocol. Because not all of the originally approved burden was utilized, this request shows all of the still needed burden but requests only the difference between the needed and unused burden amounts.

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 SBTs 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 request is to conduct the following pretesting activities related to NAEP 2017 assessments:
- Cognitive interviews with students for grade 8 reading tasks; and
- Small-scale student tryouts for grade 4 and 8 mathematics and grade 8 reading.

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;

---

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

- Appendices J-AI — recruitment and communication materials; and
- Volume II — protocols and questions used in the pretesting sessions.

Types of Pretesting

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

Cognitive Interviews
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 pretesting[2] 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 reading tasks. Using this methodology, the student's gaze is tracked as he or she works through a task, and the resulting eye movements can be interpreted to infer attentional and reasoning processes. 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 tasks)
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 Assessment[3], shows that tryout-based insights are very informative, especially for

---

[2] 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).

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

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 Digitally Based Assessments in Mathematics and Reading
In 2017, new technology-enhanced 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. Given that the assessments will be digitally based, all of the pretesting activities will be conducted using technology (e.g., a tablet or laptop)[4]. Cognitive interviews and tryouts will be conducted using draft programmed tasks.

3) Sampling and Recruitment Plans

The sampling and recruitment plans for each type of testing are described below.

Cognitive Laboratories (Reading)
ETS will conduct the cognitive interviews for reading. 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. 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 as part of exploratory cognitive interviewing.[5] A sample size of 5 to 15 individuals

---

[4] For the ease of description, the term "computer" has been used in the recruitment materials.

[5] 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:
ftp://akmc.biz/ShareSpace/ResMeth-IS-Spring2012/Zhora_el_Gauche/Reading%20Materials/Someren_et_al-The_Think_Aloud_Method.pdf

has become the standard for NAEP cognitive interviews.[6] Based on this research and prior experience, six 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 completed per session and the number of tasks to be cognitively tested, cognitive interviewing is expected to involve a maximum of 24 students for grade 8 reading[7].

Small-scale Tryouts for Mathematics and Reading

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 50 students for each scenario-based task. In addition to the SBT, students may take selected technology-enhanced discrete items. Up to 400 students will be recruited for small-scale tryouts across grades 4 and 8 for mathematics and grade 8 for reading. Students will participate in tryouts

---

[6] 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).

[7] The original package included cognitive interviews of 120 students for mathematics and reading. To date 12 students have participated in the cognitive interviews. Therefore, the total number of cognitive interviews for reading and math for the entire study decreases from 120 to 36 (12 completed + 24 to be conducted). Because only 1/10 of the burden approved for cognitive interviews was utilized, we are carrying over a credit of 666 respondents, 918 responses, and 272 burden hours.

in groups[8]. Table 1 summarizes the number of students for the, cognitive interviews and tryout components of the cognitive pretesting activities.

Table 1. Sample Size: Cognitive Pretest Activities: Cognitive Interviews, Tryouts [9]

	Grade 4	Grade 8	Grade 12	Total
Cognitive Interview	0	24	NA	24
Tryouts	100	300	NA	400
Total	**100**	**324**	**NA**	**424**

4) Data Collection Process

Cognitive Laboratories (Reading)

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 do reading 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 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

---

[8] The original package included tryouts of 200 students for mathematics and reading. To date, 150 students have participated in the tryouts. Therefore, the total number of tryouts for reading and math for the entire study increases from 200 to 550 (150 completed + 400 to be conducted). Because only 3/4 of the burden approved for tryouts was utilized, we are carrying over a credit of 348 respondents, 592 responses, and 91 burden hours.

[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 per pretesting activity will not exceed the total shown in the table.

verbalizing their thoughts. For example: "What's going on in your head right now?" and "I see you're looking at this screen. 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. Eye movements are unobtrusively recorded and all events on the screen are captured in real time.

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 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.

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).

Analysis Plans
For the cognitive interview data collections, documentation will be grouped at the task level. The types of data collected about task items and components will include:
- think-aloud verbal reports;
- behavioral data (e.g., reactions to avatars where present, errors in reading items or tasks;);
- 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)
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. In contrast to the cognitive interviews, in the tryouts there will be no think-aloud or verbal probing component, although students will be given a set of questions about their

reactions to the task and their experiences with the subject area (reading or mathematics). 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.

<u>Analysis Plan</u>

Student responses to items will be compiled to allow quantitative and descriptive analyses of the performance and a subset of process data (e.g., time on various task sections). Standard item analyses will be performed, in addition to frequency counts, along with order information and analysis of patterns of responses and some actions captured.

5) Consultations Outside the Agency

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

<u>EurekaFacts</u>, 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.

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 recorded[10]. 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.

---

[10] Recordings will be audio only, as described in the specific interview sections.

8) Estimate of Hourly Burden

Cognitive Interview Burden – Reading

The estimated burden for recruitment assumes attrition throughout the process.[11] The anticipated number of student participants for these cognitive interviews is 24 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 70 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 90 minutes for grade 8 students. Table 4 details the estimated burden for the reading cognitive laboratories.

Table 4. Estimate of Hourly Burden for Cognitive Interviews for Reading

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	64	21
Parent or Legal Guardian			
Flyer and consent form review	0.08	64	5
Consent form completion and return	0.13	32*	4
Confirmation to parent via email or letter	0.05	24*	2
Recruitment Totals		**128**	**32**
Student			
Grade 8 Student Interview	1.5	24	36
Interview Totals		**24**	**36**
Total Burden		**152**	**68**

* Subset of initial contact group (**total number of responses = 208**)

Small-Scale Tryout Burden – Mathematics and Reading

The estimated burden for recruitment assumes attrition throughout the process.[12] The anticipated number of student participants for small-scale tryouts is 400. 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.

---

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

[12] Assumptions for approximate <u>attrition rates</u> 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	2,775	139
Follow-up via phone, including consent form completion and return	0.15	1,110*,**	167
Confirmations	0.05	444**	22
Recruitment Totals		**2,775**	**327**
Student			
Grade 4	1	100**	100
Grade 8	1	300**	300
Interview Totals		**400****	**400**
Total Burden		**2,775**	**728**

* This includes both parents and students from 555 households
** Subset of initial contact group (**total number of responses = 4,729**)

Total for All Pretesting Activities

The combined totals for all of pretesting activities are listed in Table 7 (minus the unused portion of burden in the original approval – 1850-0803 v.88).

Table 7. Combined Burden for Pretesting Activities (minus the unused portion of burden in the original approval – 1850-0803 v.88).

	Number of respondents	Number of responses	Burden Hours
Cognitive items and tasks			
Total Cognitive Interview Burden	**152**	**208**	**68**
Total Tryout Burden	**2,775**	**4,729**	**727**
Overall Sub-totals	**2,927**	**4,937**	**796**
Unused Carry Over from Last Approval	**1,014**	**1,510**	**363**
Total Burden	**1,913**	**3,427**	**432**

9) Estimate of Costs for Recruiting and Paying Respondents

For all student pretesting activities held outside of school hours, a $25 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 reading cognitive interviews take place at schools during school hours, the $25 gift cards will be given to the school administrators.

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 Cognitive Interviews Design, prepare for, and conduct cognitive interviews (including recruitment, incentive costs, data collection, analysis, and reporting).	ETS	$ 242,426
Cognitive Item Small-scale Tryouts Design, prepare for, and conduct scoring and analysis of tryouts.	ETS	$ 112,250
Prepare for and conduct tryouts (including recruitment, incentive costs, data collection, reporting).	EurekaFacts	$197,066
Total		**$551,742**

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
Cognitive interviews for reading	August 2015-November 2015
Small-scale tryouts for mathematics and reading	August 2015-January 2016