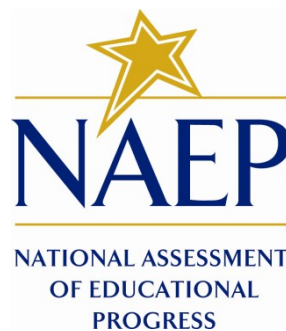


*NATIONAL CENTER FOR EDUCATION STATISTICS
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS*

*Volume I
Supporting Statement*

OMB# 1850-0803 v. 164

*National Assessment of Educational Progress (NAEP) 2019
Science Items Pretesting*



August 2016

Table of Contents

| | | |
|-----|--|----|
| 1) | Submittal-Related Information..... | 3 |
| 2) | Background and Study Rationale..... | 3 |
| 3) | Sampling and Recruitment Plan..... | 4 |
| 4) | Data Collection Process..... | 6 |
| 5) | Analysis Plan..... | 7 |
| 6) | Consultations Outside the Agency..... | 7 |
| 7) | Assurance of Confidentiality..... | 8 |
| 8) | Justification for Sensitive Questions..... | 8 |
| 9) | Estimate of Hourly Burden..... | 8 |
| 10) | Estimate of Costs for Recruiting and Paying Respondents..... | 9 |
| 11) | Costs to Federal Government..... | 10 |
| 12) | Schedule..... | 10 |

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 cognitive interviews and tryouts) to test new methodologies, question types, or delivery methods to improve survey and assessment instruments.

This submittal requests clearance for pretesting activities (cognitive interviews and tryouts) of technology-enabled discrete items and interactive computer tasks (ICTs) related to the 2019 science assessment at grades 4, 8, and 12.

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 conducted 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 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, pretesting methods such as cognitive interviews and tryouts are used to obtain data about new digitally enhanced items. These methods are intended to enhance the efficiency of the development of assessment instruments, before formal piloting. Pretesting also helps NCES identify and eliminate, as much as possible, problems with items before those items are used in formal pilots. This, in turn, means fewer challenges in scoring and analysis, higher pilot item survival rates, less revisiting of test design, and thus time efficiencies gained in operationalizing items. As NAEP moves forward with digitally based assessments, pretesting is especially important given unknown factors associated with innovative digitally based item types.

The primary focus of the technology-enabled discrete item pretesting will be to determine whether any item content or presentation causes confusion or introduces construct-irrelevant variance. ICT pretesting will also determine whether the ICTs elicit targeted knowledge and skills, as well as allow for the observation of task speed, item difficulty, and instances of missing data. Tryouts allow for the collection of information regarding whether students can finish the ICTs in the allotted time, whether students get the item correct, and where data intended to be collected during the testing process is not collected.

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;
- Appendices A-U – recruitment and communication materials;
- Appendices V-AA – screeners and consent forms; and
- Volume II – protocols and questions used in the pretesting sessions.

Types of Pretesting

Cognitive Interviews

In cognitive interviews (often referred to as cognitive laboratory studies or cog labs), an interviewer uses a structured protocol in a one-on-one interview, drawing on methods from cognitive science. In NAEP studies to date, two methods have been 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. With verbal probing techniques, the interviewer asks 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 the students to verbalize their thought processes in an unconstrained way, supplemented by specific and targeted probes from the interviewer, has proven to be flexible and productive. This will be the primary approach in forthcoming NAEP cognitive interviews.

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. The main objective is to explore how students are thinking and what reasoning processes they are using as they work through sets of items or ICTs. Along with the main interviewer, a second observer will document informal observations of student behavior and verbalizations. Behavioral observations may include nonverbal indicators that suggest emotions such as frustration or engagement, and interactions with the set of items or ICTs, such as prolonged time on one item or ineffectual or repeated actions suggesting misunderstanding.

Cognitive interviews may be conducted for draft programmed builds. The general approach will be to have a small number of participants work through the set of items or ICTs. Data will then be synthesized in the form of lessons learned about inferred student cognitive processes and observed student behaviors. The participant's performance on the set of items or ICTs will be considered on a number of levels, from basic usability issues to questions of validity, such as whether the ICT items appear to be eliciting the constructs of interest. These lessons will then inform ongoing item or ICT development.

Small-Scale Tryouts

In tryouts, students will work uninterrupted through selected set of draft programmed items or ICTs. Tryouts provide a small-scale snapshot of the range of responses and action items that are drawn out, which can be gathered much earlier in the assessment development process and with fewer resource implications than formal piloting. During try-out sessions, process data will be recorded by the delivery platform eNAEP, which will support analysis and lessons learned for ongoing item and ICT development.

Sampling and Recruitment Plan

Recruitment Plans

ETS, the item developer for NAEP, will be responsible for the overall administration of the pretesting activities described in this package. ETS and EurekaFacts, a subcontractor to ETS, will administer the cognitive interviews, and EurekaFacts will administer the tryouts.

For the pretesting sessions, students will be recruited from the following demographic populations:

- A mix of race/ethnicity (Black, Asian, White, Hispanic)
- A mix of socioeconomic background
- 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.

For the pretesting sessions administered by ETS, 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 teachers and staff at local schools and afterschool programs for students via emails or letters. Paper flyers and consent forms for students and parents will be distributed through teacher and staff contacts.

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 existing targeted mailing lists containing residential addresses and landline telephone listings. EurekaFacts will also use a participant recruitment strategy that integrates multiple out-reach/contact methods and resources such as newspaper/Internet ads, outreach to community-based organizations (e.g., Boys and Girls Clubs, Parent-Teacher Associations),

social media, and mass media recruiting (such as postings on the EurekaFacts website). To ensure that the sample population is representative of different geographical areas (urban, rural, and suburban), for the pretesting sessions administered by EurekaFacts, students will be recruited from the District of Columbia, Maryland, Virginia, Delaware, and Southern Pennsylvania.

To accommodate participants (recruited by both ETS and EurekaFacts) from different areas and to lessen their burden, pretesting session will take place at the ETS campus in Princeton, New Jersey, at EurekaFacts' Rockville (Maryland) site, and at other venues convenient to participants, such as after-school or community-based organizations. In all cases, a suitable environment (e.g., quiet rooms) will be used for the pretesting sessions, and more than one adult will be present.

Interested participants (recruited by both ETS and EurekaFacts) will be screened to ensure that they meet the criteria for participation in the pretesting session (e.g., their parents/guardians have given consent and they are from the targeted demographic groups outlined above). When recruiting participants, staff will first speak to the parent/guardian of the interested minor before starting the screening process. 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 email/letter and/or phone call. Informed consent from students (over 18 years of age) and parents (for students under 18 years of age) will be obtained for all respondents. In an effort to facilitate communication with parents or guardians of potential participants, some recruiting materials may be translated into other languages.

Sampling Plans – Cognitive Interviews

Existing research and practice have failed to offer a methodological or practical consensus regarding the minimum or optimal sample size necessary to provide valid results for cognitive interviews and similar small-scale activities.¹ Nonetheless, a sample size of five to fifteen individuals has become the standard. Several researchers have confirmed the standard of five as the minimum number of participants per group for analysis for the purposes of exploratory cognitive interviewing.²

Accordingly, seven to ten students per set of items or ICT should be sufficient given that the key purpose of the cognitive interview is to identify qualitative patterns in how students think at different points when responding to items. Given the number of items to be developed, a maximum of 78 students across grades 4, 8, and 12 will participate in cognitive interviews. The interviews will take 60 minutes for grade 4 and 90 minutes for grades 8 and 12.

Sampling Plans – Tryouts

The optimal number of students for small-scale tryouts will vary depending on the methodology in place for that specific tryout. Tryouts may include between five and 20 students. For example, a small-scale tryout ending with a short discussion portion, where the moderator asks students to discuss their thoughts and impressions on different aspects of the task, would work for groups no more than eight students. If there is a short discussion portion at the end, students tend to be more comfortable and researchers are able to gain a lot more insight and information on the tasks if there are about 5-8 students participating in a single session. Alternatively, if the tryouts do not consist of a post-tryout discussion portion with the moderator, the optimal number participating in the pretesting activity can be between 15-20 students.

A maximum of 688 students will be recruited for small-scale tryouts across grades 4, 8, and 12. The

¹ See Almond, P. J., Cameto, R., Johnstone, C. J., Laitusis, C., Lazarus, S., Nagle, K., Parker, C. E., Roach, A. T., & Sato, E. (2009). White paper: Cognitive interview methods in reading test design and development for alternate assessments based on modified academic achievement standards (AA-MAS). Dover, NH: Measured Progress and Menlo Park, CA: SRI International.

² See 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.

tryout sessions will take 60 minutes for grade 4 and 90 minutes for grades 8 and 12.

Table 1 illustrates the sample size by pretesting type and grade.

Table 1. Sample Size: Cognitive Interviews and Tryouts^{3,4}

| | Grade 4 | Grade 8 | Grade 12 | Total |
|----------------------|----------------|----------------|-----------------|--------------|
| Cognitive Interviews | 24 | 24 | 30 | 78 |
| Tryouts | 374 | 194 | 120 | 688 |
| Total | 398 | 218 | 150 | 766 |

4) Data Collection Process

Cognitive Interviews

Participants will first be welcomed by staff and introduced to the interviewer and the observer, then they will be told that they can ask any questions. 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. Depending on the content being tested, some of the cognitive interviews will utilize retrospective methods in which students complete items/task, and then review their work at the end while describing what they were thinking about at the time. A replay of the set of items/task (e.g., using software like Camtasia[®] or Morae[®]) will help cue the student’s recollection of their thought processes that occurred during the set of items/task as the student progressed through it. Other cognitive interviews will utilize concurrent methods consisting of think-aloud interviewing and verbal probing techniques. Both concurrent and retrospective methods utilize think-aloud interviewing in which students are explicitly instructed to “think aloud” (i.e., describe what they are [or were] thinking) as they work [or worked] through questions. 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.

The protocols for the think-aloud sections 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 ICT [or screen/figure/chart/text]. What are you thinking?*” The think-aloud method also includes a verbal probing component conducted after completion of the think-aloud portion for a given item or set of items. These verbal probes include a combination of pre-planned questions, identified before the session as important and ad hoc questions from in situ observations during the interview (e.g., clarifications or expansions on points raised by the student). For example, if a student paused for a long time over a particular item, appeared to be frustrated at any point, or indicated an “aha” moment, the interviewer might probe these kinds of observations further, to find out what was going on. 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 set of items. ETS will prepare the welcome script, think-aloud instructions, and hints for the interviewers (e.g., how to probe for question clarity in a task, how to probe for clarity of stimulus in a task as it leads to item presentation, and how to probe for rationale for responses to items in a task).

Additionally, ETS or EurekaFacts will administer a series of follow-up questions at the end of each cognitive interview to help identify prior knowledge, specific issues with the items or task, and level of familiarity with computers.

³ This table represents the expected distribution across grades. Depending on the nature of the items and ICTs and the specific recruitment challenges, the actual distribution may slightly vary. For burden purposes, the maximum number of students by pretesting activity will not exceed the total shown in the table.

⁴ While students in grades 4, 8, and 12 will be targeted, students in grades 5, 9, and 11 will also be allowed to participate, based on recruitment needs and challenges.

See Volume II for the specific protocols to be used in the study.

Small-Scale Tryouts

Tryout sessions will be administered by EurekaFacts in small groups. Because tryouts are sessions where the students complete items on their own without any interruption, it is most efficient to have several students complete items at the same time. In addition to the science items and/or ICTs, students will respond to follow-up questions to help identify prior knowledge, specific issues with the items or task, and level of familiarity with computers. Some tryouts will also have a proctor-lead discussion session related to task-specific issues.

A proctor will be present during the tryout sessions and will follow a strict protocol to provide students with general instructions, guide the group through the tryout, facilitate the discussion session (for the tryout sessions that have one) and assist students in the case of any technical issues. In addition, the proctor will take notes of any potential observations or issues that occur during the tryout session and include follow up questions about specific items or issues observed.

Analysis Plan

Cognitive Interviews

For the cognitive interview data collections, documentation will be grouped at the ICT, discrete item, set, or block level. Items will be analyzed across participants. The types of data collected will include

- think-aloud verbal reports;
- behavioral data (e.g., errors in reading items and actions observable from screen-capture, and gaze patterns where collected);
- responses to generic questions prompting students to think out loud;
- responses to targeted questions specific to the item(s);
- additional volunteered participant comments; and
- answers to 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 item/task components: For example, patterns of responses to probes or debriefing questions or types of actions observed from students at specific points in an ICT. 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 and provide recommendations for addressing those problems.

Small-Scale Tryouts

Student responses to discrete items will be compiled into spreadsheets to allow quantitative and descriptive analyses of the performance data. Completion times and non-completion rates will also be quantified and entered into the spreadsheets. These datasets will be used in the development, design, and programming decisions.

6) Consultations Outside the Agency

ETS, as the NAEP Item Development and the NAEP Science ICT contractor, will be responsible for all activities described in this package, including recruitment for and administration of some of the cognitive interview sessions, and guiding and overseeing the cognitive interview and tryout sessions administered by EurekaFacts. EurekaFacts owns and operates Morae[®] software, which allows for video and audio capture of students being interviewed, and to capture all of the students' moves on the computer, including mouse clicks, mouse trails, time on task, and places for interviewers to insert comments and notes at any part of the task. The software also provides remote access to video so that NCES and ETS

staff can observe the interviews from a distance in real time.

7) 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 legal guardians (of minor students) and directly from students 18 years or older before interviews are administered. Participants will be assigned a unique student 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, secured for the duration of the study, and will be destroyed after the final report is released. The interviews will be recorded using video and audio technology. The only identification included on the files will be the participant ID. The recorded files will be secured for the duration of the study and will be destroyed after the final report is submitted.

8) Justification for Sensitive Questions

Throughout the item, task, and process of developing interview protocols, effort has been made to avoid asking for information that might be considered sensitive or offensive.

9) Estimate of Hourly Burden

The estimated burden for recruitment assumes attrition throughout the process. Assumptions for approximate attrition rates for direct participant recruitment from initial contact to follow-up are 75 percent, from follow-up to confirmation 20 percent, and from confirmation to participation 20 percent. All cognitive interview and tryout sessions will be scheduled for 60 or 90 minutes. Table 2 details the estimated burden.

Table 2. Burden for Science Cognitive Interviews

| Respondent | Hours per respondent | Number of respondents | Number of responses | Total hours (rounded up) |
|---|-----------------------------|------------------------------|----------------------------|---------------------------------|
| Cognitive Interviews | | | | |
| Student Recruitment via Teachers and Staff | | | | |
| Initial contact with staff: e-mail or phone, flyer distribution, and planning | 0.33 | 15 | 15 | 5 |
| Parent or Legal Guardian for Student Recruitment | | | | |
| Initial contact | 0.05 | 492 | 492 | 25 |
| Follow-up contact | 0.15 | 123* | 123 | 19 |
| Consent form completion and return | 0.13 | 98* | 98 | 13 |
| Confirmation | 0.05 | 98* | 98 | 5 |
| Sub-Total | | 492 | 811 | 62 |
| Participation (Cognitive Interviews) | | | | |
| Grade 4 Students | 1 | 24 | 24 | 24 |
| Grade 8 Students | 1.5 | 24 | 24 | 36 |
| Grade 12 Students | 1.5 | 30 | 30 | 45 |
| Sub-Total | | 78 | 78 | 105 |
| Total Burden Cognitive Interviews | | 585 | 904 | 172 |
| Tryouts | | | | |
| Student Recruitment via Teachers and Staff | | | | |
| Initial contact with staff: e-mail or phone, flyer distribution, and planning | 0.33 | 50 | 50 | 17 |
| Parent or Legal Guardian for Student Recruitment | | | | |
| Initial contact | 0.05 | 4,300 | 4,300 | 215 |
| Follow-up contact | 0.15 | 1075* | 1,075 | 162 |
| Consent form completion and return | 0.13 | 860* | 860 | 112 |
| Confirmation | 0.05 | 860* | 860 | 43 |
| Sub-Total | | 4,300 | 7,095 | 532 |
| Participation (Tryouts) | | | | |
| Grade 4 Students | 1.5 | 374 | 374 | 561 |
| Grade 8 Students | 1.5 | 194 | 194 | 291 |
| Grade 12 Students | 1.5 | 120 | 120 | 180 |
| Sub-Total | | 688 | 688 | 1,032 |
| Total Burden Tryouts | | 5,038 | 7,833 | 1,581 |
| Total | | 5,623 | 8,737 | 1,753 |

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

10) Estimate of Costs for Recruiting and Paying Respondents

A \$25 major credit card (e.g., Visa) gift card will be offered to each student for participating in a pretesting session, and for sessions that take place outside of school (e.g., at ETS offices) a parent or legal guardian of each student will also be offered a \$25 gift card to thank them for their time and effort bringing their student to and from the pretesting site.

11) Costs to Federal Government

The estimated costs for the activities described in this package are provided in Table 3.

Table 3. Estimate of Costs

| Activity | Provider | Estimated Cost |
|---|-----------------|-----------------------|
| Design, preparation, implementation (including recruitment, incentive costs, data collection and documentation), analysis, and reporting of science item and task cognitive interviews. | ETS | \$ 239,215 |
| Preparation and implementation (including recruitment, incentive costs, data collection and documentation) of science item and task cognitive interviews. | EurekaFacts | \$ 349,197 |
| Design, preparation, and implementation of scoring and analysis of science item and task tryouts. | ETS | \$ 410,010 |
| Preparation and implementation of science item and task tryouts (including recruitment, incentive costs, data collection, reporting). | EurekaFacts | \$ 707,550 |
| Total | | \$1,705,972 |

12) Schedule

The cognitive interviews and small-scale tryouts (including recruitment, data collection, and analyses) will take place from September 2016 through April 2017, assuming a NAEP pilot in 2018 for a science assessment in 2019. The time period is broad because discrete items and ICTs will be ready for pretesting at different times throughout this window.