NATIONAL CENTER FOR EDUCATION STATISTICS NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Volume I Supporting Statement

NAEP Social Studies: Civics, Geography, and U.S. History Play Testing, Cognitive Interviews, and Tryouts

OMB# 1850-0803 v.138



TABLE OF CONTENTS

1)	Submittal-Related Information	1
•		
2)	Background and Study Rationale	1
3)	Sampling and Recruitment Plan	2
4)	Data Collection Process	∠
5)	Consultations Outside the Agency	<i>6</i>
6)	Assurance of Confidentiality	<i>6</i>
7)	Justification for Sensitive Questions	<i>6</i>
8)	Estimate of Hourly Burden	7
9)	Incentive Costs for Participants	9
10)	Costs to Federal Government	10
11)	Schedule	10

1) Submittal-Related Information

This material is being submitted under the generic National Center for Education Statistics (NCES) clearance agreement (OMB# 1850-0803), which allows for NCES to conduct various procedures (such as pilot tests, cognitive interviews, focus groups, feasibility studies, etc.) to test new methodologies, question types, or delivery methods to improve survey instruments and procedures. This request is to test new social studies tasks for upcoming assessments through cognitive interviews, play testing, and tryouts.

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 collect survey questionnaire (i.e., non-cognitive) data from students, teachers, and principals to provide context for the reporting and interpretation of assessment results.

As NAEP transitions from paper-and-pencil administrations to digital-based assessments (DBA), new technology-enhanced items and scenario-based tasks (SBTs) will be developed featuring a range of possible designs. One goal will be to capitalize on the digital-based environment to expand the constructs measured (for example, cognitive processes) to yield rich data in support of reporting goals. In 2018, social studies: civics, geography, and U.S. history are the subjects that are scheduled to transition to DBA.

A number of methods—play testing, cognitive interviews, and tryouts—will be used to obtain data about new digitally-enhanced items and scenario-based tasks throughout the development process. These methods are intended to enhance the efficiency of the development of assessment instruments by helping us to identify and eliminate, as much as possible, problems with items before formal large-scale pilot tests.

A range of pre-pilot testing tools allow tailoring the selected approach to the specific question or purpose to be addressed during different development stages. This submission describes these pretesting methods (play testing, cognitive interviews, and tryouts) and high-level plans for sampling and recruitment, data collection, and analysis for cognitive items and tasks for the 2018 social science assessments at grades 8 and 12.

Each pretesting method will typically be used at different stages of development. Play testing, when needed, will typically occur in early item and task development stages. Students will work with storyboards/wireframes or programmed versions of items and tasks. Cognitive interviews, when needed, will occur at the draft programmed task stage. Tryouts, when needed, will occur at the programmed task stage (this may run concurrently with play testing or cognitive labs). Thus, play testing, cognitive interviews, and tryouts will be happening simultaneously for different items and tasks at different stages of development. Data collection for tasks and items will occur on an ongoing and staggered basis. Recruitment efforts will also be ongoing during the pretesting window so that items and tasks can be pretested shortly after they are ready for whichever pretesting stage they will undergo.

Types of Pretesting

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

Play Testing

In play testing, an innovation 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 with one another and an observer/facilitator. Play testing will take place early in the process using wireframes (somewhat functional storyboards for items or tasks) or programmed task builds.

During play testing, students will be encouraged to talk together about items or tasks and issues they confront, while observers note reactions and potential problems with content or format. Observers will query students to draw them out, facilitate deeper reactions, or probe areas of possible confusion. Play testing will allow identification of construct-irrelevant features in tasks, such as inaccessible language in item stems 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

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 participants are thinking and what reasoning processes they are using to work through tasks. For these social studies cognitive interviews, retrospective think-aloud and verbal probing techniques will be employed to elicit student feedback.

Cognitive interviews will be conducted for social studies scenario-based tasks and items using draft programmed tasks. These processes are designed to evaluate tasks and provide validity evidence. The general approach will be to have a small number of participants work through targeted sections or whole tasks while data is gathered, focusing on how students work through tasks. Data will then be synthesized in the form of lessons learned about inferred student cognitive processes, observed student behaviors, and the performance of tasks on a number of levels, from basic usability issues to questions of validity, such as whether a task appears to be eliciting the constructs of interest. These lessons will then inform ongoing assessment development.

Eye tracking may also be used in the cognitive interview process. Using this methodology, the student's gaze is tracked as he or she works through an activity, and the resulting eye movements can be interpreted to infer attentional and reasoning processes. Eye tracking methods could be particularly beneficial for examining patterns of students' attention to and processing of non-interactive stimuli, during which no other information is being obtained from the student via button presses or other student-driven manipulations or actions in the environment (i.e., tasks during which the moment-by-moment logging of student actions will yield little direct evidence of students' cognition).

Tryouts

In tryouts, students will work uninterrupted through a selected set of draft programmed items or tasks intended for DBA social studies formal piloting in 2017. 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 naturalistic, uninterrupted item or task performance. Given that tryouts mimic actual administrations, data can be gathered about task timing. Finally, because tryouts will occur when tasks have been programmed to enable collection of process data, they supply an opportunity to begin collection of such data based on draft hypotheses about student behaviors in relation to the cognition targeted by a given task's measurement goals. Tryouts provide a small-scale snapshot of the range of responses and actions 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.

3) Sampling and Recruitment Plan

NCES has contracted Educational Testing Service (ETS), the NAEP cognitive item developer to carry out the pretesting activity described in this package. ETS and EurekaFacts, an ETS sub-contractor, will recruit participants and conduct the pretesting. ETS will recruit participants and conduct play testing (and possibly cog labs ¹), while Eureka Facts will recruit and conduct cog labs and tryouts.

¹ ETS may conduct some cognitive interviews, based on needs, as the pretesting activities proceed.

Interested participants will be screened to ensure that they meet the criteria for participation in the pretesting activity (e.g., their parents/guardians have given consent, they are from the targeted demographic groups, etc.). When recruiting participants, <ETS, EurekaFacts> staff will first speak to the parent/guardian of the interested minor (or the over age 18 participant) 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 email/letter. Informed parental consent will be obtained for all respondents who are interested in participating in the data collection efforts. See appendices for screeners and consent form documents.

Play Testing

ETS will recruit students from a range of demographic groups. Students will be recruited from districts that are located near the ETS campus in Princeton, New Jersey for scheduling efficiency and flexibility. Students may participate in play testing sessions only after receipt of written consent forms from their parents or legal guardians.

ETS will recruit students using existing ETS contacts with teachers and staff at local schools and afterschool programs for students. Email, letters, or phone calls will be used to contact these teachers/staff, and paper flyers and consent forms for students and parents will be distributed through these teachers/staff. 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 emails and/or letters will be sent to participants. Only after ETS has obtained written consent from the parent/guardian will the student be allowed to participate in the play testing session. Appendices A-K provide sample recruitment materials that will be used by ETS².

Five students will be convened per grade for each task; five students per grade should be sufficient at the play testing stage given that the key purpose is to identify usability errors and other construct-irrelevant issues. Based on prior experience with similar studies, it is anticipated that the same students will return to participate in multiple sessions. Therefore, play testing is expected to involve a minimum of 20 and maximum of 60 students across the grades and subjects.

Cognitive Interviews

For the cognitive interviews, students will be recruited by EurekaFacts (or ETS) staff from the following demographic populations:

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

If recruited by ETS, students will be recruited from districts that are located near the Princeton, New Jersey ETS campus for scheduling efficiency and flexibility. EurekaFacts will perform the recruiting for cognitive interviews from the District of Columbia, Maryland, Virginia, Delaware, and Southern Pennsylvania. EurekaFacts plans to conduct interviews in other venues beside their Rockville, MD site, such as after-school activities organizations or community-based organizations. This will allow them to accommodate participants recruited from other areas than Rockville, MD and ensure that the sample population is representative of different geographical areas (urban, rural, and suburban). This will also lessen participants' burden. In all cases, a suitable environment such as a quiet room will be used to conduct the interviews and there will be more than one adult present. Appendices L-AK provide sample materials that will be used for the cognitive interview recruitments.

² Note: If appropriate, relevant appendices (i.e., parental screening calls) may be translated to facilitate communication.

³ See Nielson, J. (1994). Estimating the number of subjects needed for a think aloud test. Int 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

Seven to ten students per task should be sufficient at this stage given that the key purpose of the cognitive interview is to identify qualitative patterns in how students think at different points in tasks and confirm the validity of the assessments. Based on the number of tasks that can be completed per session and the expected number of tasks to go through the cognitive interview process (up to two tasks per subject, for a total of six tasks across the social studies), cognitive interviewing is expected to involve a minimum of 42 students and maximum of 60 students across grades 8 and 12.

Tryouts

EurekaFacts will perform the recruiting for tryouts in a similar manner to their cognitive interview recruitment. Recruitment efforts will ensure that the results are representative of various populations, and specifically, inclusive of students from rural areas. As with the other types of pretesting activities, students will be sampled to obtain a mix of race/ethnicity, socioeconomic background, and location elements. Appendices AL-AY provide sample materials that will be used for the tryout recruitments.

EurekaFacts will recruit 25 students for each scenario-based task. The same students may take selected innovative discrete items that can be tested along with tasks. A total of twelve small-scale tryout sessions of 25 students each have been budgeted. A maximum of 300 students will be recruited for small-scale tryouts across grades 8 and 12.

4) Data Collection Process

The various pretesting activities will take place at ETS, EurekaFacts, or other suitable venues (e.g., a school library, after-school activities offices, community-based organizations, etc.). In all cases, a suitable environment such as a quiet room will be used to conduct the interviews and there will be more than one adult present.

Participants will first be welcomed, introduced to the interviewer and the observer, and told that they are there to help answer questions about how students respond to social studies tasks. Students will be reassured that their participation is voluntary and that their answers may be used only for research purposes. See Volume II for the protocols for the various pretesting activities.

Play Testing

Assessment specialists will give an overview of the tasks and/or items to students and provide guidance on what they should reflect on while looking at the tasks and/or items. Assessment specialists and other staff (e.g., cognitive scientists or task designers) from ETS will act as facilitators and observers, taking notes on what students say and interjecting occasional questions aimed at eliciting students' reactions, places of confusion, and ways of thinking about the answers to the questions in the tasks and/or items. Each observer may choose to stay with one group of 2–3 students looking at and responding to tasks and/or items, or they may choose to move around to observe several groups of students.

For the most part, students will be allowed to explore and interact with the mocked-up or programmed tasks and items by themselves with little intrusion on the part of the interviewer. However, at a few strategic points, the interviewer may introduce questions meant to explore students' reactions to the task, such as:

- 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?

Prior to each play testing session, interviewers may identify some key focus areas for each task. If students do not provide sufficient comments on targeted parts, a staff member 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. Student feedback from a play testing session is immediate and can be evaluated after

the session. Those items or tasks can then proceed with development with little interruption. Sessions will be audio recorded.

Analysis Plan -The results will be compiled to identify patterns of responses for tasks, including patterns of responses to probes or debriefing questions, or types of actions observed from students at specific points in composing a response to a task. This approach will help to ensure that the data are analyzed in a way that is thorough and systematic, and enhance identification of problems with tasks and developing recommendations for addressing these problems.

Cognitive Interviews

The welcome script, think-aloud instructions, and hints for the interviewers will be prepared by ETS and conducted by EurekaFacts (or ETS, if applicable). The protocols (see Volume II) 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 task [or screen/figure/chart/text]. What are you thinking?"

On completion of a task or set of items, the interviewer will proceed with follow-up questions. In this verbal probing component, the interviewer asks student targeted questions about specific aspects of knowledge, skill, or ability that the task or items are 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 a task. The targeted questions will be generated by ETS 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 an 'aha' moment, 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., Morae® software by TechSmith). Morae Recorder's core strength is its facility for capturing 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. These recordings can be replayed for later analysis, to see how a given student progressed through the task. Digital audio recording will capture students' verbal responses to the think-aloud interview, using either the tablet's integral microphone or an external digital recorder, depending on the specific tablet platform used and compatibility with the screen-capture software. 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 Plan - For the cognitive interview data collections, documentation will be grouped at the task or discrete item 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; gaze patterns where collected);
- 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
- answers to debriefing questions.

Tryouts

Tryout sessions will be conducted by EurekaFacts in small groups. Because tryouts are sessions where the students complete the task on their own without any interruption, verbal probing, or think-aloud component, it is possible and most efficient to have several students complete the task at the same time. A proctor will be present during the

session and will follow a strict protocol to provide students with general instructions, guide the group through the tryout, administer debriefing questions 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.

Analysis Plan - The focus of tryout data is particularly on score data and time to complete task and items, so the analysis will reflect these goals. Student responses to 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 data sets will be shared across staff to facilitate task development, design, and programming decisions. It will take approximately two to four weeks for the team to analyze the information and make recommendations for item, task, and scoring criteria revisions.

In addition to the final report based on all pretesting activities (mentioned in the overview section above), ETS will prepare and share with NCES a summary presentation of key findings that drive item and task revisions.

5) Consultations Outside the Agency

ETS is working with NCES to develop cognitive and survey items for NAEP assessments and is responsible for carrying out the social studies pretesting study. Its sub-contractor, EurekaFacts, is a research and consulting firm in Rockville, Maryland that offers facilities, tools, and staff to collect and analyze both qualitative and quantitative data. EurekaFacts will be involved in recruitment and the conduct of cognitive interviews and tryouts.

6) Assurance of Confidentiality

Students taking part in the pretesting activities 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 participants (over age 18) and from parents or legal guardians of students below age 18. Participants will be assigned a unique identifier (ID), which will be created solely for data file management and to keep all participant materials together, and 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 interview files, secured for the duration of the study, and destroyed after the final report is completed. The interviews will be recorded. 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 destroyed after the final report is submitted.

7) Justification for Sensitive Questions

This study does not include sensitive questions.

8) Estimate of Hourly Burden

Play Testing

The estimated burden for recruitment assumes attrition throughout the process.⁴ The anticipated maximum number of student participants for play testing is 60 (while students may participate in multiple sessions, we estimate burden based on different students participating in each session). Initial contact is estimated at 3 minutes (0.05 hours), follow-up and flyer distribution is estimated at 9 minutes (0.15 hours). We anticipate distributing 360 flyers via these contacts to parents and students. Time to review flyers is estimated at 5 minutes (0.08 hours). For filling out the consent form, the estimated time is 8 minutes (0.13 hours). The follow-up email or letter to confirm participation for each session is estimated at 3 minutes (0.05 hours). Play testing sessions are expected to last 60 minutes per student. Table 1 details the estimated burden for play testing.

⁴ Assumptions for approximate attrition rates are 75 percent from initial contact to follow-up contact and 20 percent from follow-up to confirmation/consent form completion. Note: for play testing, the school follow-up and confirmation numbers are the same.

Table 1. Burden for Social Studies Play Testing

Respondent	Number of respondents	Number of responses	Hours per respondent	Total hours		
Schools and Organizations						
Initial contact	40	40	0.05	2		
Follow-up contact/flyer dist.	10*	10	0.15	2		
Confirmation	10*	10	0.05	1		
Sub-Total	40	60		5		
Parent or Legal Guardian for St	udent Recruitment					
Initial contact/flyer review	360	360	0.08	29		
Follow-up contact	90*	90	0.15	14		
Consent form completion and return	72*	72	0.13	10		
Confirmation	72*	72	0.05	4		
Sub-Total	360	594		57		
Student Participation						
Grade 8 th	30°	30	1	30		
Grade 12 th	30°	30	1	30		
Sub-Total	60 a	60		60		
Total Burden	460	714		122		

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

Cognitive Interviews

The estimated burden for cognitive interview recruitment assumes the same attrition throughout the process as noted in play testing. All cognitive interview sessions will be scheduled for no more than 90 minutes. Table 2 details the estimated burden for cognitive interviewing.

Table 2. Burden for Social Studies Cognitive interviews

Respondent	Number of respondents	Number of responses	Hours per respondent	Total hours	
Schools and Organizations					
Initial contact	40	40	0.05	2	
Follow-up contact/flyer dist.	10*	10	0.15	2	
Confirmation	8*	8	0.05	1	
Sub-Total	40	58		5	
Parent or Legal Guardian for Stud	ent Recruitment		•		
Initial contact	376	376	0.05	19	
Follow-up contact	94*	94	0.15	14	
Consent form completion and return	75*	75	0.13	10	
Confirmation	75*	75	0.05	4	
Sub-Total	376	620		47	
Student Participation					
Grade 8	30ª	30	1.5	45	
Grade 12	30°	30	1.5	45	
Sub-Total	60°	60		90	
Total Burden	476	738		142	

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

^a Estimated number of actual participants will be somewhat less than confirmation numbers.

^a Estimated number of actual participants will be somewhat less than confirmation numbers.

Small-Scale Tryouts

The estimated burden for tryout recruitment assumes the same attrition throughout the process as noted in previous sections. All tryout sessions will be scheduled for no more than 60 minutes. Table 3 details the estimated burden for the social studies small-scale tryouts.

Table 3. Burden for Social Science Tryouts

Respondent	Number of	Number of	Hours per	Total hours	
	respondents	responses	respondent		
Schools and Organizations	Schools and Organizations				
Initial contact	40	40	0.05	2	
Follow-up contact	10*	10	0.15	2	
Confirmation	8*	8	0.05	1	
Sub-Total	40	58		5	
Parent or Legal Guardian for Stud	lent Recruitment				
Initial contact	1,875	1,875	0.05	94	
Follow-up contact	469*	469	0.15	70	
Consent form completion and return	375*	375	0.13	49	
Confirmation	375*	375	0.05	19	
Sub-Total	1,875	3,094		232	
Student Participation					
Grade 8	150°	150	1	150	
Grade 12	150°	150	1	150	
Sub-Total	300 ^a	300		300	
Total Burden	2,215	3,452		537	

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

Total for All Pretesting Activities

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

Table 4. Combined Burden for Pretesting Activities

	Number of	Number of responses	
Pretest Activity Component	respondents		Burden Hours
Total Play Testing Burden	460	714	122
Total Cognitive Interview Burden	476	738	142
Total Tryout Burden	2,215	3,452	537
Overall Totals	3,151	4,904	801

9) Incentive Costs for Participants

To encourage participation and thank them for their time and effort, a \$25 credit card gift card will be offered to each participating student. If a parent or legal guardian brings their student to and from the testing site they will also receive a \$25 gift card along with a thank you letter for allowing their child to participate in the study.

10) Costs to Federal Government

The estimated cost to federal government for the social studies cognitive interview activities is \$ 1,449,536. Table 5 shows further details of the cost.

^a Estimated number of actual participants will be somewhat less than confirmation numbers.

Table 5. Estimate of Costs to Federal Government

Activity	Provider	Cost
Design, prepare, and conduct social studies play testing activities (including recruitment, allocation of incentive costs, data collection, analysis, and reporting)	ETS	\$ 249,694
Design, analysis, and reporting for social studies cognitive interviews	ETS	\$ 216,080
Conduct social studies cognitive interviews (including recruitment, allocation of incentive costs, data collection, and reporting)	EurekaFacts	\$ 367,921
Design, prepare, and score social studies tryouts (including analysis, and reporting)	ETS	\$ 277,128
Conduct social studies tryouts (including recruitment, allocation of incentive costs, data collection)	EurekaFacts	\$ 338,713
Total Estimate		\$ 1,449,536

11) Schedule

The following high-level schedule assumes a pilot test in 2017 for a social studies assessment in 2018 at grades 8 and 12.

Table 6. Social Studies Pretesting Timeline

Activity	Dates	
Each activity includes recruitment, data collection, and analyses		
Play testing for social studies	July 2015-January 2016	
Cognitive interviews for social studies	December 2015-April 2016	
Small-scale tryouts for social studies	December 2015-April 2016	