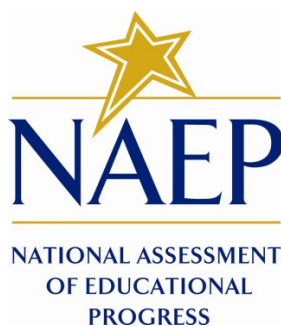NATIONAL CENTER FOR EDUCATION STATISTICS
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Volume I
Supporting Statement

The National Assessment of Educational Progress (NAEP) 2021 Items Pretesting

OMB# 1850-0803 v.219

November 2017

Contents

1. Submittal-Related Information

This material is being submitted under the National Center for Education Statistics (NCES) generic clearance agreement (OMB# 1850-0803), which provides for NCES to conduct various procedures (e.g., focus groups, cognitive interviews, usability tests, experiments, etc.) to develop and test study materials and methodologies so as to improve future data quality, utility, and study participant experience.

2. Background and Study Rationale

The National Assessment of Educational Progress (NAEP) is a federally authorized survey (by the National Assessment of Educational Progress Authorization Act; 20 U.S.C. §9622) 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 of Education Sciences, within the U.S. Department of Education. NAEP's primary purpose is to assess student achievement in different subject areas and to collect questionnaire (i.e., non-cognitive) data from students, teachers, and school administrators[1] to provide context for the reporting and interpretation of assessment results.

This request is to pretest, as part of the NAEP assessment development process, newly developed questionnaire items and assessment discrete items (DI) for the NAEP 2021 grades 4, 8, and 12 reading, mathematics, and writing[2] assessments. This pretesting, designed to enhance the efficiency of developing assessment instruments before piloting them, will utilize cognitive interviews and tryouts to identify and eliminate potential issues with new digitally enhanced NAEP items, tasks, and stimuli. Subsequently, the results should minimize challenges in item scoring and analysis and lead to higher pilot item survival rates.

The overall goal of this pretesting is to determine whether items and tasks appear to elicit targeted knowledge and skills and/or reduce construct irrelevance in the form of either evidence that can be scored or qualitative data consisting of student responses and reactions. This pretesting is designed to help identify whether any item content or features cause confusion or introduce sources of construct-irrelevant variance. The results will inform refinement of items and scoring rubrics and determination of which items will be piloted for future use NAEP. The following two types of pretesting methods will be used in this study:

1. Cognitive Interviews:

   In cognitive interviews, an interviewer uses a structured protocol in a one-on-one setting drawing on methods from cognitive science. In NAEP studies to date, two techniques 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 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* or during the process as being of particular interest. Verbal probing is also the method of choice for many cognitive interview pretesting studies for reading and writing items, because experience has shown that students have trouble articulating their thoughts while they are trying to read and comprehend texts or while they are trying to compose extended responses. For subjects or items not requiring extended reading or writing, the 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 both flexible and effective.

   Cognitive interview studies produce largely qualitative data in the form of verbalizations made by students during the think-aloud phase and/or in response to interviewer probes, using both concurrent and retrospective approaches.[3] In concurrent interviews, students are asked to verbalize the process as they progress through an item. In retrospective interviews, upon completion of each item, the interviewer proceeds with follow-up questions to collect information about students' thinking processes (e.g., "Can you tell me, in

---

[1] Please note that in this submission "school administrator" refers to the principal or assistant/vice principal. In NAEP main study administrations, other individuals who are not the head principal are allowed to complete the school administrator questionnaire.

[2] For the purposes of this document, we will reference NAEP writing tasks as discrete items.

[3] Ericsson, K.A. & Simon, H.A. (1980). Verbal reports as data. *Psychological review*, 87(3), 215- 251.
  Forsyth, B. H., & Lessler, J. T. (1991). Cognitive laboratory methods: A taxonomy. *Measurement errors in surveys*, 393-418.

your own words, what you needed to do to answer the question?"). The main objective of cognitive interviews is to explore how students are thinking and what reasoning processes they are using as they work through items and tasks. Some informal observations of behavior and verbalizations are also gathered, including nonverbal indicators of affect, suggesting emotional states such as frustration or engagement, and interactions with tasks, such as prolonged time on one item or ineffectual or repeated actions suggesting misunderstanding.

2. <u>Tryouts</u>:

   In tryouts, students will work uninterrupted through selected sets of draft items. Tryouts provide a snapshot of the range of responses and actions that items elicit, which can be gathered much earlier in the assessment development process and with fewer resource implications than piloting. Tryouts allow for pretesting of a wide range of content and the collection of more robust data on ranges of student responses, item difficulty, assessment timing, and item functionality than is practical for cognitive interviews. The larger samples and timed testing conditions of tryouts are especially useful for gathering quantitative data about items, investigating the possible effects of different item features on student performance, and learning how long it takes students to complete items. Tryout samples used to date in NAEP have ranged from smaller (50 students per item/task) to considerably larger (several hundred or more students per item/task), depending on the nature of the items/tasks, and the amount of time and resources available. For example, we have conducted small-scale tryouts to better understand item properties such as how long it takes a sample of students to complete a given item, and large-scale tryouts to compare different versions of the same item or group of items (as has been done for reading blocks based on digital passages and reading scenario-based tasks (SBTs) with and without avatars).

There are two categories of items that will be pretested as part of this study:

1. <u>Cognitive Items</u>: These items cover subject-specific knowledge in the areas of grade 4 and 8 reading, mathematics, and writing**.** Cognitive items will be part of both the cognitive interview and tryout phases of pretesting. This pretesting informs the test development process by answering the following questions:
   - Do students demonstrate a general understanding of the item and what they are asked to do, including content being measured?
   - Can students use the proposed interactive item components (IIC) integrated into discrete items?
   - How do students use the interactive item components?
   - Do student experiences differ based on their prior use of digital devices and specific tools like the interactive item components they use?
   - Do the items elicit the expected responses?
   - How do items and scoring guidelines perform, in terms of item difficulty, discrimination, and timing, in conditions similar to the final operational conditions?

2. <u>Questionnaire Items</u>: The 4th, 8th, and 12th grade core, reading, mathematics, and writing questionnaires aim to capture data related to important subject-specific and nonsubject-specific (core) contextual factors for student achievement. Questionnaires are administered to students, teachers, and school administrators. Table 1 contains the possible areas of focus for the development for upcoming NAEP questionnaires.

Table 1. Questionnaire Core Modules and Subject-Specific Areas of Focus

Core	For Each: Reading, Mathematics, and Writing
Socioeconomic Status (SES)	Resources for Learning and Instruction
Technology Use	Organization and Instruction
Perseverance	Teacher Preparation
Enjoyment of Difficult Problems	Student Factors
School Climate	N/A

Questionnaire items will be part of the cognitive interview phase of pretesting, which will inform the questionnaire development process by:
   - Identifying problems with the items (i.e., ensure the item is understood by all participants and confirming items are not sensitive in nature or make the participant uncomfortable); and

- Finding ways to improve wording of existing items where possible.

3. SAMPLING AND RECRUITMENT PLANS

Recruitment Overview

EurekaFacts and ETS will recruit participants and administer the pretesting (see Table 2). EurekaFacts will recruit for cognitive interviews (except for reading cognitive items) and tryouts from the District of Columbia, Maryland, Virginia, Delaware, and Southern Pennsylvania. ETS will recruit for reading cognitive interviews from Princeton and surrounding areas, including Trenton. EurekaFacts and ETS may administer pretesting activities at their offices in Rockville, MD and Princeton, NJ respectively, and, to obtain a sample population including different geographical areas (urban, suburban, and rural), both will likely administer interviews in venues such as after-school activities organizations and community-based organizations.

Table 2. Organization for Recruiting and Administering Pretesting Activities

Pretesting Activity			EurekaFacts	ETS
Cognitive Interviews	Cognitive Items	Reading		x
		Mathematics	x	
		Writing	x	
	Questionnaires		x	
Tryouts	Cognitive Items	Reading	x	
		Mathematics	x	

While EurekaFacts and ETS will use various outreach methods to recruit students to participate, the bulk of the recruitment will be administered by telephone. Various resources will be employed to recruit participants. For students[4] these will include:

- existing participant databases;
- targeted telephone and mail contact lists (i.e., lists that consist of individuals meeting basic criteria such as age or school grade);
- school system research/assessment directors;
- NAEP State Coordinators when possible to recruit in schools (see section 5);
- community organizations (e.g., boys/girls clubs, parent-teacher associations, and limited on-site location-based and mass media recruiting); and
- outreach/contact methods and resources (e.g., internet ads, flyers/bookmarks, canvassing, and having representatives available to talk to parents, educators, and community members at appropriate local community events, school fairs, etc.).

For questionnaires, teachers and school administrators will be recruited using the following recruitment resources in addition to those mentioned above:

- national organizations' databases of administrators and faculty;
- NCES school database (e.g., Common Core of Data and Private School Universe Survey); and
- contacts within organizations and groups that can serve as recruitment partners (e.g., Horton's Kids, Housing Authority of the City of Frederick) and, if needed, targeted contact lists.

A general overview of the recruitment process for the pretesting activities is as follows:

1) EurekaFacts or ETS will send an email of introduction about the pretesting research to: (a) various elementary, middle, and high school principals (to recruit students); (b) individuals in the subcontractors' existing databases; (c) community centers/organizations and research/assessment directors; (d) targeted telephone and mail/email contact lists; (e) parents/guardians; and (f) teachers and principals (Appendices B-F).[5] The email of introduction will include an informational brochure (Appendix U).

---

[4] For students under age 18, parents/legal guardians will receive the various contact information.

[5] Note that in 1(a) principals are being targeted to identify schools where students, teachers, and/or school administrators could be recruited, while in 1(f) are principals from the EurekaFacts databases that are being recruited specifically for the school administrator interviews.

2) EurekaFacts or ETS will only discuss additional recruitment materials, such as flyers and informational bookmarks (Appendices T and V), with those community organizations that contact EurekaFacts/ETS upon receiving the email of introduction and informational brochure.

3) After receiving a contact of interest, a EurekaFacts or ETS staff member will follow up with the interested parent/legal guardian of an under-age student, student age 18 or older, teacher, or school administrator via phone and ask them to provide demographic information to ensure that a diverse sample is selected (Appendices W-Z).

4) If the parent/legal guardian allows their student to participate, or the student age 18 or older, teacher, or school administrator agree to participate, EurekaFacts or ETS will follow up to confirm participation and the date and time of the cognitive interview session (Appendices K-L).

5) Parents/legal guardians (on behalf of the students under 18), students age 18 or older, teachers, and school administrators will be required to sign consent forms prior to the pretesting session (Appendices AB-AD).

EurekaFacts and ETS will recruit 4th, 8th, and 12th grade students (a mix of gender, race/ethnicity, English language learner status, and socioeconomic background), teachers (a mix of school sizes and school demographics), and school administrators (a mix of school sizes and school demographics) so that a diverse sample is achieved. Please note that SES will be given the highest priority during recruitment while also ensuring sufficient balance of the other criteria. The subcontractor will document the information collected in the screeners using a tracking sheet, which will be used to determine the targeted sample, including diversification on key characteristics (see Appendix AH for example tracking sheet). Additionally, it should be noted that the sample is not large enough to support subgroup analyses.

To minimize the travel burden on students, parents/legal guardians, teachers, and school administrators, cognitive interviews will be administered in nearby venues that are convenient for the participants, such as the EurekaFacts offices in Rockville, MD, community centers, facilities of community-based organizations, and school building sites (after school only). All student cognitive interviews and the majority of teacher and school administrator cognitive interviews will be administered in-person.[6]

Sampling

Existing research and practice have not identified 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.[7] Nonetheless, a sample size of 5 to 15 individuals has become the standard for NAEP studies.

Cognitive Items Cognitive Interviews:

Based on the research, seven to 10 participants per block or set of items will be sufficient given that the key purpose of the cognitive interview is to identify qualitative patterns in how students are reasoning at different points during or after responding. Total number of participants by subject area is as follows:

Table 3. Approximate Sample Size for Cognitive Items Cognitive Interviews*

Subject	Number of Students Grade 4	Number of Students Grade 8
Reading	10	10
Mathematics	10	10
Writing	60	30

*Note that numbers may be reallocated based on results of early pretesting.

Questionnaire Cognitive Interviews:

Table 4 summarizes the number and types of cognitive interviews that are planned to test both core and subject-specific questionnaire items. Note that a minimum number of five respondents per subgroup is recommended to

---

[6] If needed, a limited number of teacher/administrator interviews may be administered via phone or WebEx.

[7] 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. Available at: http://www.measuredprogress.org/documents/10157/18820/cognitiveinterviewmethods.pdf.

identify major problems with an item and for a meaningful analysis of data to test the usability of developed prototype questions.[8] Students and teachers will be oversampled to better ensure detection of items that may cause confusion or raise sensitivity issues.

Table 4. Approximate Sample Size for Questionnaire Cognitive Interviews[9]

Respondent Group	Grade 4	Grade 8	Grade 12	Grades 4/8/12	Total
Students	20	20	10	N/A	**50**
Teachers	10	10	N/A	N/A	**20**
School Administrators (Non-Charter School)	5	5	5	N/A	**15**
School Administrators (Charter School)	N/A	N/A	N/A	5	**5**
Overall Total	**35**	**35**	**15**	**5**	**90**

No more than three students will be recruited per school. No more than one teacher or school administrator will be recruited per school.

Cognitive Items Tryouts:

Tryouts are intended to reflect naturalistic test-taking conditions that allow for larger samples than cognitive interviews in order to survey a representative range of completion times and responses. Table 5 provides the approximate sample sizes that will be used for tryouts.

Table 5. Approximate Sample Size of Student Participants for Tryouts*

Subject	Number of Students Grade 4	Number of Students Grade 8
Reading	150	100
Mathematics	50	50
Writing**	N/A	N/A
Overall Total	**200**	**150**

*Note that numbers may be reallocated based on results of early pretesting.
**There are no tryouts planned for writing.

4. DATA COLLECTION PROCESS

Data Collection for Cognitive Interviews:

Participants will first be welcomed, introduced to the interviewer and the observer (if an in-room observer is present), and told that they are there to help us ensure that students/teachers/administrators like them understand the newly developed core and subject-specific items (see Volume II). Participants will be reassured that all of the information they provide may be used only for statistical purposes (see section 6). As part of the introduction process, the interviewer will explain to participants that their responses will be audio recorded. For the phone/web-based teacher and school administrator cognitive interviews, the interviewer will explain the technology and describe the tools the participants may use, such as muting their phone and asking questions.

The interviewer will be tasked with keeping participants engaged by asking probe questions (see Volume II), soliciting responses from less talkative participants, and asking follow-up questions where appropriate (e.g., "That's interesting, could you tell me a little bit more about that?"). Interviewers may also take additional notes during the in-person cognitive interviews, including noting observed behaviors (e.g., the participant's facial expressions indicated they are confused) and whether extra time was needed to answer certain questions. Please refer to Volume II for the specific protocols and item probes for the various survey questions being pretested. ETS or

---

[8] 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.

[9] Grade 4 and 8 students and teacher participants will receive core and mathematics items, or reading and writing items. Grade 12 students will receive core and writing items. School administrators (non-charter school) will receive items for all subjects. School administrators from charter schools will first receive the charter-school specific items and, if time allows, will be administered the other non-charter school specific items.

EurekaFacts staff may record audio and screen activity for analysis.

Cognitive interviews for cognitive items can last up to 60 minutes for grade 4 and 90 minutes for grade 8. Cognitive interviews for questionnaires can last up to 90 minutes[10] for students, teachers, and school administrators.

Analysis for Cognitive Interviews:

The types of data collected about items will include:
- think-aloud verbal reports;
- process data (e.g., timing);
- behavioral data (e.g., signs of frustration or interest; actions observable from interviewer notes);
- responses to generic questions prompting students to think aloud;
- 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: for example, patterns of responses to probes or debriefing questions, or types of actions observed from students at specific points. 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.

Data Collection for Tryouts:

Tryout sessions will be administered in small groups. Because during tryouts students complete items on their own without any interruption, it is possible and most efficient to have several students work 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 any debriefing questions, and assist students in the case of any technical issues. The proctor will take notes of any potential observations or issues that occur during the tryout session. Finally, it may be desirable once students have completed their work, and time allowing, for proctors to present students with follow-up verbal or written probes, typically asking students about their reactions, areas of confusion, and background knowledge (see Volume II).

The tryout sessions will be scheduled for 60 minutes at grade 4 and 90 minutes at grade 8. Reading tryouts will be a combination of NAEP reading blocks and of a brief measure of students' reading skills (e.g., the Gates-MacGinitie Reading Test). Mathematics tryouts will only include NAEP mathematics blocks. Process data will be collected along with student responses. A small set of survey questions may be used to collect additional debriefing and/or engagement information from students (see Volume II). ETS or EurekaFacts staff may record audio and screen activity for analysis.

Analysis for Cognitive Items Tryouts:

The types of data collected will include:
- responses to items;
- process data (e.g., timing and students' movements among items, opening and closing of the item panel in a reading DI block)
- EurekaFacts observer notes; and
- answers to any debriefing questions.

Student responses to items will be compiled into spreadsheets to allow quantitative and qualitative 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 assessment instruments development.

---

[10] Please note that the 90 minutes includes time for introductions (maximum 15 minutes), conducting the interview (60 minutes), and debriefing and/or time for additional questions/feedback from the participants (maximum 15 minutes).

5. CONSULTATIONS OUTSIDE THE AGENCY

ETS serves as the Item Development, Data Analysis, and Reporting contractor on the NAEP project, developing cognitive and survey items for NAEP assessments. As such, ETS will be responsible for the managing the administration of all activities described in this package. Additionally, ETS will recruit and administer the cognitive interviews for reading cognitive items. EurekaFacts, a subcontractor for ETS, will recruit and administer the cognitive interviews and tryouts for cognitive items subject areas and questionnaires. The NAEP State Coordinators serve as the liaisons between state education agencies and NAEP, coordinating NAEP activities within their respective states. NAEP State Coordinators from selected states may provide leads for potential participants for this study.

6. ASSURANCE OF CONFIDENTIALITY

The study will not retain any personally identifiable information. Prior to the start of the study, participants will be notified that their participation is voluntary. As part of the study, participants will be notified that all of the information they provide may be used only for statistical purposes and may not be disclosed, or used, in identifiable form for any other purpose except as required by law (20 U.S.C. §9573 and 6 U.S.C. §151).

Written consent will be obtained from parents or legal guardians of students under the age 18, or from students who are age 18 or older at the time of participation. Verbal assent will be obtained from all participating students under the age of 18. Participants will be assigned a unique identifier (ID) at the time of recruitment, which will be created solely for data file management and used to keep all participant materials together. The participant ID will be separated from the participant's name before the report from each stage of the study is finalized. The signed consent form, which includes the participant name, will be separated from the participant interview files. The interviews may be audio recorded. The only identification included on the audio files will be the participant ID. All consent forms, recordings, and individual records/notes will be secured for the duration of the study and will be destroyed after the final report is completed

7. JUSTIFICATION FOR SENSITIVE QUESTIONS

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

8. PAYING RESPONDENTS

To encourage participation, a $25 gift card from a major credit card company will be offered to each student who participates in each pretesting session as a thank you for his/her time and effort. For sessions that take place in locations other than schools, a parent or legal guardian of each student will also be offered a $25 gift card from a major credit card company to thank them for bringing his/her participating student to and from the testing site. Teachers and school administrators who participate in a questionnaire pretesting session will be offered a $40 gift card from a major credit card company as a thank you for their time and effort.

9. ESTIMATE OF HOURLY BURDEN

The estimated burden for recruitment assumes attrition throughout the process. Cognitive interviews and tryout sessions for cognitive items are expected to take 60 minutes for grade 4 students and 90 minutes for grade 8 students. Cognitive interviews for questionnaires are expected to take 90 minutes in all cases.

Table 6. Estimate of Hourly Burden for Cognitive Interviews and Tryouts

Respondent	Number of respondents	Number of responses	Hours per respondent	Total hours
Principal/School Administrator or Point Person for Community Organizations for Student Recruitment				
Initial contact	129	129	0.05	7
Follow-up & Identify students	85*	85*	1.0	85
Sub-Total	**129**	**214**	**-**	**92**
Parent or Legal Guardian for Student Recruitment				
Initial contact	1,492	1,492	0.05	75
Follow-up via phone	781*	781*	0.15	117
Consent & Confirmation	550*	550*	0.15	83
Sub-Total	**1,492**	**2,823**	**-**	**275**
Teacher and School Administrator Recruitment				
Initial contact	135	135	0.05	7
Follow-up via phone or e-mail	100*	100*	0.15	15
Consent & Confirmation	50*	50*	0.15	8
Sub-Total	**135**	**285**	**-**	**30**
Participation (Cognitive Interviews for Questionnaires)**				
Students	50	50	1.5	75
Teachers	20*	20*	1.5	30
School Administrators	20*	20*	1.5	30
Sub-Total	**50**	**90**	**-**	**135**
Participation (Cognitive Interviews for Cognitive Items)				
Grade 4 (Mathematics, Reading, Writing)	80	80	1	80
Grade 8 (Mathematics, Reading, Writing)	50	50	1.5	75
Sub-Total	**130**	**130**	**-**	**155**
Participation (Tryouts for Cognitive Items)				
Grade 4 (Reading and Mathematics)	200	200	1	200
Grade 8 (Mathematics)	50	50	1	50
Grade 8 (Reading)	100	100	1.5	150
Sub-Total	**350**	**350**	**-**	**400**
Total Burden	**2,286**	**3,892**	**-**	**1,087**

* Subset of initial contact group

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

10. COST TO FEDERAL GOVERNMENT

Table 7. Total Estimated of Costs of Pretesting

Activity	Provider	Estimated Cost
Prepare and administer cognitive interviews and tryouts (including recruitment, incentive costs, data collection, and documentation for all pretesting activities except for cognitive reading item cognitive interviews)	EurekaFacts	$1,253,000
Design, prepare for, conduct scoring and analysis, and prepare report (including recruitment, incentive costs, and data collection for cognitive reading items)	ETS	$432,000
	Total	**$1,685,000**

11. PROJECT SCHEDULE

Recruitment for each form of pretesting will begin in November 2017, upon OMB approval. Data collection and analyses for questionnaire cognitive interviews are scheduled to end in February 2018 and for cognitive items cognitive interviews and tryouts in March 2018.