Developing Achievement Levels on the National Assessment of Educational Progress for Writing at Grade 4 Design Document

Submitted: December 6, 2016

Submitted to:
National Assessment Governing Board
800 North Capitol Street, NW, Suite 825
Washington, DC 20002-4233

This study was funded by the National Assessment Governing Board under Contract ED-NAG-16-C-0001.

Submitted by:
NCS Pearson, Inc.
2510 N. Dodge Street
Iowa City, IA 52245-9945
Phone: 319.354.9200
# Table of Contents

Executive Summary ................................................................................................................. 4

Section 1: Achievement Levels-Setting Methodology .............................................................. 8

Section 2: Achievement Levels Panels .................................................................................. 10
  Identification of Panelist Nominators .................................................................................. 12
  Selection of Panelists .......................................................................................................... 14

Section 3: Briefing Materials ................................................................................................ 15

Section 4: Field Trial .............................................................................................................. 16
  Materials ............................................................................................................................. 19
  Participants .......................................................................................................................... 20
  Procedures .......................................................................................................................... 21
  Results ............................................................................................................................... 24

Section 5: Pilot Study ............................................................................................................. 25

Section 6: Achievement Levels-Setting Tasks and Procedures ............................................ 26
  Preparing for the ALS Meeting .......................................................................................... 27
  Preparing Panelists to make ALS Judgments ...................................................................... 31
  Collecting Content-Based Judgments .................................................................................. 33

Section 7: Information Processing and Feedback ................................................................. 38

Section 8: Identifying Potential Exemplar Student Responses ............................................ 44

Section 9: Public Comment .................................................................................................. 44

Section 10: Validation Studies .............................................................................................. 47
Examination of Procedural Validity ................................................................. 47
Examination of Internal Validity ................................................................. 48
Examination of External Validity ............................................................... 50
References ........................................................................................................ 53
The National Assessment of Educational Progress (NAEP), known as the “The Nation’s Report Card,” provides information on what students in the United States know and can do in various subject areas.

The 2017 grade 4 NAEP writing assessment is the first administration of the grade 4 assessment developed to meet the design specifications described in the current computer-based Writing Framework (https://www.nagb.org/publications/frameworks/writing/2017-writing-framework.html)\(^1\). The assessment is a digital-based assessment, comprised of constructed response items, for which students compose and construct their responses using word processing software on a tablet. The test administration software incorporates commonly available word processing tools. Pursuant to the Governing Board’s legislative mandate, achievement levels must be set for the grade 4 writing assessment. In accordance with the Governing Board policy on setting achievement levels for NAEP, the achievement levels-setting process includes achievement level descriptions (ALDs), cut scores, and exemplar items. In 2012, the Governing Board formally approved the updated achievement level descriptions for writing at all three grade levels that describe what students should know and be able to do at three levels: Basic, Proficient, and Advanced.

The writing assessment at grade 4 will be administered in January to March 2017 to a nationally representative sample of approximately 22,000 grade 4 students. The assessment is designed to assess three communicative purposes: to persuade, to explain, and to convey experience. In addition, student responses will be evaluated on three broad features of writing: development of ideas, organization of ideas, and language facility and conventions. For grade 4, the audience for the students’ writing will be specified or clearly implied and will include such audiences as peers, teachers, and parents. In addition,\(^1\)

\(^{1}\) The first administration of the grade 8 and grade 12 NAEP writing assessment under this framework occurred in 2011, with achievement levels-setting in 2012.
grade 4 students will be assigned a specific form (e.g., short story or newspaper editorial) to use in responding to the writing prompt. For more information on the design of the NAEP writing assessment, please visit http://nces.ed.gov/nationsreportcard/writing/whatmeasure.aspx.

Achievement levels have become a powerful way to communicate student achievement on an assessment like the NAEP writing because achievement levels interpret test performance with reference to cut scores that quantitatively define ordered categories of achievement such as basic or proficient (Haertel & Lorie, 2004). An important source of evidence used by policymakers to establish achievement levels is the cut score recommendations that result from an achievement levels-setting (ALS) meeting. Cut scores are the outcome of a facilitated process, called a standard setting meeting, that systematically elicits judgments from experts related to the test content and the knowledge, skills and abilities of the test takers (Hambleton, Pitoniak & Copella, 2012).

The National Assessment Governing Board has issued a contract to Pearson to implement a process to produce a set of cut score recommendations to assist the National Assessment Governing Board in developing achievement levels for the 2017 grade 4 NAEP writing assessment. On behalf of the Governing Board, Pearson has developed this Design Document that describes in detail the grade 4 NAEP writing achievement levels-setting activities. This document is intended to provide the foundation for all ALS activities. The Design Document will be used to guide all aspects of the achievement levels-setting process that, in the end, produce a set of cut score recommendations for reporting achievement levels for the 2017 administration of the grade 4 NAEP writing.

For standard setting, Pearson proposed using the Body of Work (BoW) standard-setting approach (Kingston, Kahl, Sweeney, & Bay, 2001; Kingston & Tiemann, 2012) for which panelists will make content-based cut score recommendations. The BoW methodology is a holistic standard setting method for which panelists evaluate sets of examinee work (i.e., bodies of work) and provide a holistic judgment about each student set. The BoW methodology is ideally suited to assessments with extended constructed response items, such as the NAEP writing assessments comprised of two constructed
response prompts. Panelists review student responses to two writing prompts, and holistically categorize
the body of responses for each student into a performance level. These categorizations are used to
compute cut scores. The panelists’ content-based judgments will be made over three rounds. The process
to be implemented for the grade 4 NAEP writing achievement levels-setting meetings (field trial, pilot
study, and operational ALS meeting) follows BoW procedures used in 2011 for NAEP writing standard
setting studies at grades 8 and 12. The field trial will be completed prior to the pilot study to provide an
opportunity to try out a number of key aspects of the ALS plan, including the logistical design of the
ALS studies such as the use of tablets and laptop computers, the ease with which the panelists can enter
judgments and questionnaire responses, and the arrangement of tables and panelists.

In alignment with the 2017 grade 4 NAEP writing administration, as well as the process
implemented for grades 8 and 12 in 2011, Pearson will use computers during the field trial, pilot study,
and operational ALS meeting. Using computers and an online interface housed in Moodle, an open-
source online e-learning platform, will reduce the time required for panelists to complete most steps in
the standard setting activities. The use of computers strengthens the validity argument for the ALS
outcomes because panelists will experience the assessments in the same context as the students
assessed and have a greater capacity to accurately evaluate the cognitive demands the assessment
imposed on grade 4 students.

Below is a summary of what each section in the Design Document includes.

- Section 1: Achievement Levels-Setting Methodology focuses on the rationale for and
  the application of the Body of Work (BoW) standard-setting approach (Kingston, Kahl,
  Sweeney, & Bay, 2001; Kingston & Tiemann, 2012) to set achievement levels for
  writing at grade 4.

- Section 2: Achievement Levels Panels describes the panelist identification and
  recruitment plan designed to obtain broadly representative and well-qualified panelist
groups for all studies.
• Section 3: Briefing Materials describes the briefing materials sent to panelist prior to each panel study in the ALS process (field trial, pilot study, and operational ALS).

• Section 4: Field Trial describes the use of a field trial to provide an opportunity to try out a number of key aspects of the ALS plan focused on logistics and implementation of hardware and software.

• Section 5: Pilot Study describes the pilot study designed to incorporate lessons learned from the field trial and to use the exact procedures planned for the operational ALS meeting. The only difference planned between the pilot study and the operational ALS is the number of panelists.

• Section 6: Achievement Levels-Setting Tasks and Procedures describes the ALS tasks, the nature of the tasks, and the procedures to be implemented prior to and as a part of the operational ALS meeting—including how panelists are trained and supported in implementing all activities.

• Section 7: Information Processing and Feedback describes the use of technology (hardware and software); provides samples of the feedback Pearson plans to use to inform panelists about their individual and group judgments; and describes the analysis plan for data obtained from all ALS studies.

• Section 8: Identifying Potential Exemplar Items explains the rationale and process for identifying potential exemplar items to recommend to the Governing Board to illustrate performance on NAEP writing at grade 4.

• Section 9: Public Comment delineates the strategies planned for collecting and using public comment regarding the Design Document and the ALS outcomes.

• Section 10: Validation Studies describes the planned approach for collecting and evaluating procedural, internal, and external validity evidence.
Performance standards have become a powerful way to communicate student achievement because they interpret test performance quantitatively, with reference to cut scores, by defining ordered categories such as basic or proficient (Haertel and Lorié, 2004). An important source of evidence that policymakers like the National Assessment Governing Board (hereafter referred to as the Governing Board) use to establish performance standards is the cut score recommendations that result from an Achievement Levels-Setting (ALS) meeting. The first step along the path to developing cut score recommendations is the choice of a standard setting method appropriate to the tasks posed to panelists and preparations for implementing the chosen method with panelists (Hambleton, Pitoniak, & Copella, 2012).

**Selecting a Standard Setting Methodology**

The ALS methodology used for the grade 4 NAEP writing assessment must meet all requirements for NAEP ALS as described in the Governing Board policy on developing student achievement levels for NAEP (https://www.nagb.org/content/nagb/assets/documents/policies/developing-student-performance.pdf) and be appropriate for the NAEP writing assessment that is comprised of two prompts and results in a measure of student performance based on scoring of a student’s response to these prompts. Three achievement levels-setting procedures have been designed and implemented for NAEP writing since 1992. Pearson carefully considered the research conducted and lessons learned from previous panel studies for NAEP writing ALS projects and designed the grade 4 writing ALS process to address the fidelity of writing ALS results across grades by closely following procedures successfully implemented for grades 8 and 12 (Measured Progress, 2012a). As such, Pearson proposed using the

---

2 Results of the 1992 Achievement Levels-Setting process were not reported; see American College Testing (1992). Setting achievement levels on the 1992 National Assessment of Educational Progress in reading, mathematics, and writing: Technical Report for further information.
Body of Work (BoW) standard-setting approach (Kington, Kahl, Sweeney, & Bay, 2001; Kingston & Tiemann, 2012) to set achievement levels for NAEP writing at grade 4.

The BoW methodology was selected for two key reasons. First, as noted above, it is the same methodology used to set achievement levels for the NAEP writing assessments at grades 8 and 12 in 2011 (Measured Progress, 2012a). Using the same approach for grade 4 writing will provide consistency across the ALS procedures implemented for the NAEP writing assessments and removes the potential for differences in the achievement level cut scores due to the use of different standard setting methods. Second, the BoW methodology is ideally suited to assessments with extended constructed-response items, such as the NAEP writing assessments comprised of two constructed-response prompts, and is intuitive for panelists to implement. Panelists review student responses to two writing prompts, and holistically categorize the body of responses for each student into a performance level. These categorizations are used to compute cut scores. The BoW approach also satisfies the main considerations for choosing an appropriate standard setting methodology (Hambleton & Pitoniak, 2006): (a) the method is appropriate for the item types and item scaling, (b) the judgments are likely to be completed in a reasonable amount of time, (c) the Governing Board has experience with the standard setting method, and (d) the measurement field appears to view the method as supported by current validity evidence.

Given that the 2017 grade 4 NAEP writing assessment will be administered for the first time using tablets with detachable keyboards, panelists are likely to have concerns regarding the consistency, across the nation, of grade 4 students’ experience and skills with using tablets for a writing assessment. To address these potential panelist concerns, Pearson will summarize and share the information collected by the NAEP program that explains the elements considered in the decision to use tablets for the grade 4 writing assessment, and the results of the analysis of pilot study data reported by the National Center for Education Statistics (http://nces.ed.gov/nationsreportcard/writing/lessons/performance.aspx). While the information
provided from these sources is not likely to eliminate all concerns, it provides an empirical and factual basis for panelists to consider when making their judgments. Pearson plans to share this information with field trial, pilot study, and operational ALS meeting panels.

The ALS process should reflect the emphasis on technology incorporated into the administration of the assessment. Pearson plans to use technology in as many aspects of the ALS procedures as possible. To support panelists’ understanding of the context of the grade 4 NAEP assessment, panelists will take the assessment using the same tablets that students will use for the 2017 assessment. In addition, panelists will be assigned a laptop computer and will use Moodle throughout the standard setting process. Pearson anticipates that using a Moodle platform on laptop computers to implement the BoW methodology during both the pilot study and the operational ALS meeting will reduce the time required for panelists to complete most steps in the standard setting process, minimize data entry errors by having panelists directly enter their judgments, and support efficient and effective sharing of feedback. The use of this platform will be evaluated through the field trial. Pearson will establish a secure Moodle website that panelists can log into to access all non-secure materials prior to the meeting, that can be used with meeting specific logins during the meeting, and that facilitates panelists’ data entry of judgments as well as their review of and interaction with feedback. Panelists will access Moodle in advance of their panel meeting to review non-secure briefing materials. The Moodle platform is described further in Section 7 of this document.

Section 2: Achievement Levels Panels

Pearson will implement a multi-step panelist recruitment plan for the field trial and, separately, for the pilot study and operational ALS meeting. Panelists for the field trial will be recruited using the same process as that applied to recruitment of the pilot study and operational ALS panels, with the exception of geographic location of panelists. The field trial panelists will be recruited from the geographic area in proximity to the field trial site, whereas the pilot study and operational ALS panels will be recruited from across the nation. Both recruitment efforts are designed to obtain
broadly representative, well-qualified panelist groups whose members are familiar with the
to knowledge, skills, and abilities grade 4 students need for writing. Pearson will recruit a total of 75
panelists—20 for the field trial, 22 for the pilot study, 33 for the operational ALS meeting. Panels
will reflect an overall balance of gender, race/ethnicity, geographic location (pilot study and
operational ALS only), and public versus private schools. Classroom teachers currently engaged in
writing instruction at grade 4 will compose 55 percent of each panel. Other educators, composing
about 15 percent of each panel, will include individuals such as higher education faculty teaching
elementary education courses, librarians, and state and local English language arts curriculum
directors. Representatives of non-educator groups will compose approximately 30 percent of each
panel and will be identified based on their background or experience in writing as well as with grade
4 students.

Each panelist pool will include at least one teacher of English Language Learners and one
teacher of students in Special Education programs; these teachers must also meet the general teacher
panelist qualifications. As noted above, Pearson will recruit 22 and 33 panelists, respectively, for the
pilot study and operational ALS meeting to ensure a full panel (a minimum of 20 for the pilot study
and 30 for operational ALS meeting) in the event that one or more panelists have to drop out before
the panel meeting.

For the field trial, panelists will be drawn from a local pool of educators and non-educators who
live within 30 miles of the field trial facility. This may limit Pearson’s ability to obtain the broad
representation intended on other panel characteristics, however, every effort will be made to ensure the
representativeness and quality of the field trial panelists. For the pilot study and operational ALS
meeting, panelists will be drawn from across all regions of the country and balanced to represent each
region (northeast, south, midwest, and west). The phases of each recruitment process include:

Phase 1: Identify nominators through organizations (see description below) and other pertinent
sources; contact nominators and ask them to nominate qualified educators and non-educators via an
online nomination form (e.g., name, contact information, and basic qualifications). For the field trial, this will be limited to members of organizations familiar with the writing educator and non-educator pools in the area surrounding San Antonio, Texas. Nominators will be asked to indicate if they judge the nominee to be “outstanding” in terms of their professional performance and to briefly describe the rationale for that judgment when given.

Phase 2: Notify nominees; request résumés and completed online nomination form (e.g., background in writing instruction, professional achievements, experience with grade 4 students). Nominees will be asked to specify training and experience that makes them an outstanding candidate for panel selection.

Phase 3: Evaluate nominated candidates based on their background and experience; select the most qualified panelists and assign to panel groups with respect to gender, race/ethnicity, type of writing instruction experience, and type of institutional affiliation.

Pearson will use email and phone calls to communicate with nominators and nominees. An honorarium of $500 will be paid to panelists. In addition, expenses in compliance with federal travel regulations will be reimbursed. For the field trial this will include expenses for travel (i.e., mileage) while for the pilot study and operational ALS meeting this will include travel (i.e., airfare), lodging, and meals. Substitute teacher costs will be reimbursed directly to schools based on actual school costs for substitute teacher payments.

Identification of Panelist Nominators

Panelist nominators will be recruited using multiple sources. One source for nominators will be the organizations that were involved in the development of the NAEP grade 4 Writing Framework, provided feedback on the framework, or have a strong background in providing professional development in writing education. The following organizations are among those that will be targeted for recruiting panelists:

- National Writing Project
• National Council of Teachers of English
• The National Center for Literacy Education
• National Education Association
• American Federation of Teachers
• Parent Teacher Association

In addition to these organizations, state superintendents, heads of teacher organizations (state and local levels for the field trial and state level for the pilot study and operational ALS meeting), school board presidents, and district and school administrators of public and private elementary education entities will be contacted in the four NAEP regions to propose qualified nominators across all three panelist types (teacher, non-teacher educator, and non-educator). Based on previous recruitment efforts (Measured Progress, 2012a; Pearson, 2016), the target ratio for public to private school nominators will be 9:1. The process of recruiting panelists for the non-teacher educator group will also include contacting professionals with connections to the authors of children’s books, deans of a representative sample of teacher education programs at higher education institutions, and NAEP state coordinators.

Based on previous experiences in recruiting NAEP ALS panelists, (ACT, 2010; Measured Progress, 2012a; Pearson, 2016), Pearson estimates that 20 percent of the nominators will respond by submitting at least one nominee for consideration. We further estimate that no more than 20 percent of the nominees would meet the qualifications, satisfy the requirements for representation, and agree to serve on the panel. Thus, for the field trial Pearson estimates that 500 nominators must be identified to yield at least 100 active nominators and at least 100 nominees. Assuming that 20 percent of those nominees will be eligible, meet the distribution requirements for representation on the panel, and be available/agree to serve as panelists, the yield would be the target of 20 panelists. Pearson will supplement the number of nominators, as needed, to attain the field trial panel targets. For the pilot study and operational ALS recruitment, Pearson estimates that 1375 nominators must be identified to yield at least 275 active nominators and at least 275 nominees. Assuming that 20 percent of those
nominees will be eligible, meet the distribution requirements for representation on the panels, and be available/agree to serve as panelists, the yield would be the target of 55 panelists. Again here, Pearson will supplement the number of nominators, as needed, to attain the panel targets.

**Selection of Panelists**

Nominees will be asked to complete an online questionnaire regarding their qualifications and experiences for serving on the panel. Candidates that present with the credentials required of panelists will be contacted by phone to collect any missing information, verify the information provided, and confirm the willingness of the candidate to serve on the panel if selected. The goal is to select the most qualified panelists who are knowledgeable about grade 4 writing, while maintaining the goal to recruit 55 percent teachers, 15 percent non-teacher educators, and 30 percent non-educators to compose each of the panels. Panelists nominated in each panelist group must meet the following qualifications:

**Teacher panelist:**
- At least five years of overall teaching experience, and
- At least two years of experience teaching writing at grade 4, and
- Judged to be “outstanding” in their professional performance by a nominator

**Non-teacher educator panelist:**
- Non-teacher educational staff at elementary schools with education and/or experience in grade 4 writing, or
- Curriculum director or content specialist serving elementary level schools or state department of education with education and/or experience in grade 4 writing, or
- Postsecondary teacher education faculty teaching courses in elementary writing instruction

**Non-educator panelist:**
- Person with expertise in writing who is not a former educator, and
- Familiar with students in grade four (e.g., as a parent or volunteer)
The credentials of panelists will be evaluated and scored based on the number and importance of the credentials that are presented. Nominees having no distinguishing credentials will score low. Nominees having extensive credentials, including having been named outstanding teacher/teacher of the year and/or being actively engaged at the national level in professional activities within the area of writing, will score high. The scoring scheme differs for each panelist type (teacher, non-teacher educator, and non-educator). Nominees with the highest scores are given top priority by placing the best-qualified candidates at the beginning of the candidate list. The selection process then selects panelists to reach the targets for representation listed above, with nominees having the highest qualifications being the first selected each time. All panels will be selected to have approximately equal proportions of males and females and equal proportions of representation from each of the four NAEP regions. We will also attempt to draw panels so that at least 20 percent of the panelists self-identify as a minority.

As noted above, in addition to covering the direct expenses for panelists (consistent with federal travel regulations), panelists for the field trial, pilot study, and operational ALS meeting will be given an honorarium of $500 each to cover incidental expenses during their stay at the panel meetings. Pearson acknowledges that the funds available to offer panelists are not commensurate with their contribution. They will emphasize that panelists’ participation in the grade 4 NAEP writing ALS represents an exceptional contribution to writing education in the United States.

Section 3: Briefing Materials

Pearson will send access to a set of briefing materials to each confirmed panelist for their review and familiarization prior to the relevant panel meeting (field trial, pilot study, or operational ALS). The Moodle platform site, customized specifically for the project, will provide panelists with secure online access. The first time the panelist logs in to the website, he or she must read and electronically sign a nondisclosure agreement. Once signed, the panelist will be guided through a brief online training for using the Moodle platform before having access to the non-secure advanced materials.
designated for the particular panel study through links on the Moodle site. Panelists will use an online checklist to guide them through the online materials and will be able to check off each document after it has been reviewed. Documents will include the following:

- Confidentiality agreement
- Purpose of the meeting
- Meeting roles and responsibilities
- Request for reimbursement form
- Meeting agenda
- NAEP Writing Framework
- NAEP Writing ALDs
- Briefing Booklet
- Hotel information, including directions
- Other materials identified as appropriate

Communication with panelists will encourage engagement with the briefing materials as they are intended to serve as a foundation for successfully carrying out the process designed for each panel. Pearson staff will be able to check the Moodle interface to determine which materials a panelist has accessed.

Section 4: Field Trial

Pearson plans to use technology in multiple ways throughout the ALS process (see Section 7 of this document for more detail). Having panelists experience the grade 4 assessment on the tablets used by students allows them to more accurately judge the cognitive demands placed on the students under the same conditions students experienced when taking the test. As previously noted, Pearson anticipates that using a Moodle platform on laptop computers to implement the BoW methodology during both the pilot study and the operational ALS meeting will reduce the time required for panelists to complete most
steps in the standard setting process, minimize data entry errors by having panelists directly enter their judgments, and support efficient and effective sharing of feedback.

Pearson will conduct a two-day field trial at the Pearson facilities in San Antonio, Texas on June 5-6, 2017 to try out key aspects of the ALS plan. Prior to the field trial Pearson will develop the Moodle site, allowing for its implementation and evaluation as part of the field trial process. All aspects of the Moodle site design will be developed by the end of January 2017, followed by the incorporation of NAEP grade 4 items and statistics needed to support the further development of the Moodle site in preparation for the ALS activities. Pearson will then revise and refine the Moodle site based on the results of (1) a detailed quality control process, (2) review of the Moodle site and its functionality by the COR and TACSS, and (3) initial and final internal testing. All revisions will be in place prior to the field trial in June 2017.

The field trial will focus on testing aspects of the logistical design of the ALS procedures such as panelists’ use of tablets and laptop computers, and the ease with which panelists use the Moodle software platform to enter judgments, respond to questionnaires, and evaluate feedback. The TACSS will be consulted for input to the study design prior to the field trial. Since the spring 2017 operational data for the NAEP grade 4 writing assessment will not be available until September 2017, all field trial data (e.g., student response data) will be based on data from the 2012 grade 4 NAEP writing pilot administration. The writing prompts and rubrics will be those used for the operational administration in 2017. These are the same prompts and rubrics used for the pilot with minor, non-substantive revisions made between the pilot and operational administrations.

The field trial will address the following questions:

- NAEP assessment:
  - Does the use of the tablets for experiencing the grade 4 NAEP writing assessment pose any difficulties for panelists?
What training do panelists need to make effective and efficient use of the tablets when taking the NAEP assessment?

- **ALS procedures:**
  - Does the room set-up support effective interactions among panelists and support panelists’ work throughout the ALS process?
  - Does the Moodle interface function effectively for panelists? More specifically, does it function effectively for each of the following:
    - Review of briefing materials,
    - Review of training materials,
    - Review of ALDs,
    - Review and rating of BoWs,
    - Transfer of ratings from paper form to Moodle,
    - Interpretation and use of feedback, and
    - Completion of process evaluation and consequences questionnaires?
  - Are the ratings and BoW review comments captured and transferred accurately:
    - By panelists when transferring information from paper to Moodle interface, and
    - By the system when capturing and transferring information from Moodle to the analysis tool?
  - Are questionnaire responses captured and transferred accurately from Moodle to the analysis tool?
  - Is feedback created after each round of standard setting transferred accurately from the analysis tool to Moodle and displayed correctly through the Moodle interface?
  - Are panelists able to use the Moodle Interface efficiently and effectively?
Did panelists have difficulty staying within the time allotted for each activity?

**Materials**

The setup and materials used for the field trial will mirror those planned for use within the grade 4 NAEP writing pilot study and operational ALS meeting. This includes the computers, the BoWs, and the type of briefing and training materials used within the study.

The sample of BoWs selected for use in the study will be selected from the student responses to the pilot conducted in 2012. The procedures implemented for the field trial will include activities that represent each major component of the full ALS activity planned for the pilot study and operational ALS meeting, and will include two sets of BoWs with both overlapping and unique BoWs within each set. Specifically, the groups of BoWs used for the study will be selected according to the criteria used in the process implemented for writing at grades 8 and 12 in 2011 (Measured Progress, 2012b). There are 22 unique writing tasks at grade 4 and each task is specific to a purpose for writing—to convey, to explain, or to persuade. From the 22 tasks, 44 forms were created for the 2017 grade 4 NAEP writing assessment, with each task appearing on four forms—twice as a first task and twice as a second task. Each writing assessment form is composed of two tasks for different purposes, with tasks using four different stimuli—image, text, audio, and video.

As was implemented for the 2011 writing NAEP ALS for grades 8 and 12, Pearson will use all writing tasks, but not all forms, for setting achievement levels on the grade 4 writing assessment. They will select a subset of eleven forms, ensuring that all 22 tasks are represented on the subset of forms. The 20 panelists recruited for the field trial will be divided into two sets of 10 to minimize the load on each panelist. As such, Pearson will assign seven forms to each panelist subgroup, three forms will be common across the two subgroups and four will be unique to a subgroup. The common forms support evaluation of the consistency of results. The common forms will include at least one of each type of prompt. Pearson will assure that the average difficulty (i.e., mean score) for the tasks is approximately equal across the tasks addressed by the two subgroups. The assignment of forms will ensure each group
is assigned the same number of tasks for each writing purpose and the same average difficulty within each writing purpose. See the description in Section 6 for further detail on selection and assignment of forms and tasks.

BoWs (each comprised of two responses by the same student) will then be selected from the forms identified for use in the field trial. Using the identified forms, Pearson will select a total of 78 BoWs from the subset of student responses to the 2012 pilot study for which the student received a total raw score of 2-12, with each response scored 1-6, and neither response coded ‘non-scorable,’ ‘off-task,’ or ‘blank.’ We will use a stratified sampling process to select seven BoWs from each form with the exception of one form from which eight BoWs will be selected. BoWs will be selected to ensure a uniform distribution across the scale of expected a posteriori (EAP) ability estimates calculated for individual BoWs and that there is not a concentration of BoWs from a given form in any score range. Panelists in each subgroup will review 50 bodies of work, 22 of which will be common across both subgroups. Based on findings of the studies conducted in 2011 for NAEP writing at grades 8 and 12 (Measured Progress, 2012a), the BoWs will be presented in descending order according to student performance based on the 2012 grade 4 NAEP pilot study. See the description in Section 6 for further detail on selection and review of BoWs.

**Participants**

Pearson will recruit panelists for the field trial as described in Section 2 of this document. Again, the focus is to create well-qualified panels comprised of panelists who are familiar with the knowledge, skills, and abilities for writing at grade four. A total of 20 panelists—70 percent educators and 30 percent non-educators—will be recruited for the field trial. Panelists will reflect an overall balance of gender, race/ethnicity, type of writing experience, and type of institutional affiliation. Classroom teachers currently engaged in writing instruction at grade 4 will compose 55 percent of each panel. Other educators, composing about 15 percent of each panel, will include individuals such as higher education faculty teaching elementary education courses, librarians, and state and local English language
arts curriculum directors. Representatives of non-educator groups will be identified based on their background or experience in writing as well as with grade 4 students.

**Procedures**

The field trial will take place in a Pearson facility in San Antonio, Texas. The room selected for the field trial will allow Pearson to use a room set-up similar to that planned for the pilot study and operational ALS meeting. Participants will be asked to review a packet of materials prior to the field trial to help familiarize them with the ALS activities. These materials will be provided through the Moodle platform. Prior to engaging with the pre-meeting materials, participants will complete a brief online training on how to navigate and use the Moodle platform.

The pre-meeting panelist briefing materials will include:

- Confidentiality agreement
- Purpose of the meeting
- Meeting roles and responsibilities
- Request for reimbursement form
- Meeting agenda
- NAEP Writing Framework
- NAEP Writing ALDs
- Briefing Booklet
- Meeting information, including directions
- Other materials identified as appropriate

Field trial activities will be completed across the two days allotted for the study. The facilitator will lead the field trial meeting and guide panelist through a series of activities designed to mirror each major component of the pilot study and operational ALS meeting. Pearson staff will implement a validation process to check the accuracy of information when panelists transfer it from paper to electronic form as well as when it is transferred between Moodle and the tools used for analysis.
Field trial activities will include:

- Information about grade 4 students’ tablet use:
  - Pearson will provide a summary of the information collected by the Governing Board that explains the elements considered in their decision to use tablets for the grade 4 writing assessment as well as the results of the analysis of pilot study data reported by the National Center for Education Statistics (http://nces.ed.gov/nationsreportcard/writing/lessons/performance.aspx).
  - For the field trial only: Before tablet research is shared, panelists will respond to a survey question asking about grade 4 students’ use of tablets to assess their achievement in writing. Panelists will respond to the same question after the research on tablet use is shared. This question will include an open-ended component, asking panelists to explain their response—including the impact of the evidence presented on that response.
  - The facilitator will lead a short discussion of panelists’ concerns if, after reviewing the relevant research, panelists continue to believe grade 4 students cannot use a tablet for writing.

- Process overview during which the facilitator will explain the objectives for the field trial, review the agenda, and provide a brief explanation of each step in the field trial procedures.

- Testing functionality of the tablets for panelists to experience taking the grade 4 NAEP writing assessment. This portion of the field trial includes:
  - Panelists logging into and taking the grade 4 NAEP writing assessment on the tablets students will use for the 2017 assessment administration.
  - Facilitator providing training on the grade 4 writing rubrics and application of the rubrics by panelists to their responses on the assessment.
  - Facilitator will lead a discussion with panelists that focuses on their experience
when taking the grade 4 assessment. Pearson staff will take notes throughout this
discussion for summarization and use in conjunction with other information
collected throughout the field trial.

- Implementing Moodle training. This portion of the field trial includes:
  - Facilitator providing additional training, if needed, on the Moodle interface.
  - Panelists navigating the Moodle interface and responding to a questionnaire on their
    understanding of and comfort level in navigating the Moodle interface.
  - Follow-up to address questions and concerns.

- Panelist training on and review of the grade 4 writing ALDs. This portion of the field trial
  includes:
  - Facilitator providing an abbreviated orientation to and training on the grade 4
    writing ALDs.
  - Panelist review and discussion of the grade 4 writing ALDs.
  - Brief questionnaire focused on panelists’ perceptions of this activity and its utility
    within the ALS procedures.

- Panelist training on and implementation of the BoW methodology. This portion of the field
  trial includes:
  - Facilitator providing training on the BoW method for standard setting.
  - Panelist completion of a readiness form, followed by a practice exercise.
  - Panelist completion of one round of ratings of the BoWs selected for the field trial
    and their review of the feedback that will be provided across the three rounds of
    ratings used for the full ALS process implemented for the pilot study and the
    operational ALS meeting.
    - Panelist review and rating of bodies of work using the Moodle interface.
    - Panelist completion of pre and post round process questionnaires.
- Pearson analysis of ratings and process data; and provision of rating round feedback to panelists including consequences data.
- Panelist completion of consequences data evaluation and completion of consequences questionnaire.

- Panelist training on and implementation of exemplar performance selection. This step in the process includes:
  - Facilitator providing abbreviated training on the selection of exemplars (see Section 8 of this document for more information regarding the exemplar selection process).
  - Panelist selection of exemplars for at least one achievement level.
  - Panelist completion of process questionnaire.

- Facilitator will lead a discussion focused on the panelists’ use of the Moodle interface and on their implementation of the BoW methodology. Pearson staff will take notes throughout this discussion for summarization and use in conjunction with other information collected throughout the field trial.

**Results**

Pearson will analyze all data collected during the field trial and provide a summary of findings to the COR, COSDAM, and the TACSS. Data will primarily include panelist responses to all surveys and questionnaires across every functional activity listed above, from pre-meeting engagement to selection of exemplars. Surveys will be designed to capture ratings according to Likert-type rating scales ranging from Strongly Agree to Strongly Disagree. Summaries of these data will include frequency counts of each rating by survey question as well as means and standard deviations. Additional summaries will be provided for all open-ended survey questions, notes taken during a meeting, and recorded observations. Based on results of the field trial and recommendations based on these results, Pearson will identify and resolve any process-related or logistical issues prior to the pilot study.
Section 5: Pilot Study

Pearson will conduct a three-and-a-half-day pilot study in Atlanta, Georgia from November 6-9, 2017 to incorporate lessons learned from the field trial and to implement the exact procedure planned for the operational ALS meeting. Conducting the pilot study at this time offers an opportunity to preview, revise, and resolve issues prior to the operational ALS implementation in February of 2018. To maintain uniformity of conditions, the pilot study and operational ALS meeting will be held at the same venue with the same agenda of activities, and all steps in the pilot study will be the same as those planned for the operational ALS meeting. The only difference planned between the pilot study and the operational ALS is the number of panelists. Like the field trial, the pilot study will utilize 20 panelists rather than the 30 recommended for the operational ALS meeting.

The TACSS will be consulted for advice and recommendations regarding details of the design prior to and after the pilot study, and up to two members of the TACSS will be asked to observe the pilot study. All pilot study materials will be based on data from the 2017 grade 4 NAEP writing operational administration. The pilot study has the following goals:

- Determine whether modifications for training, instructions, materials, timing, and logistics will be needed for the operational ALS meeting, and
- Provide an opportunity for facilitators to practice the process before moving to the operational ALS meeting.

Given that all steps in the pilot study represent those planned for the operational ALS meeting, details on the ALS process for both the pilot study and the operational ALS meeting are provided under Section 6 of this document which clearly describes the operational ALS meeting.

A pilot study report will be prepared, no later than December 19, 2017, for presentation to COSDAM during a webinar or conference call. The timing of the pilot study allows for preparation of the report and review of the report by the COR, COSDAM and TACSS in advance of the operational ALS meeting. The Governing Board staff will consider their observations and all feedback from the
Section 6: Achievement Levels-Setting Tasks and Procedures

Pearson will conduct a three-and-a-half-day operational ALS meeting in Atlanta, Georgia February 12-15, 2018. To maintain uniformity of conditions, the operational ALS meeting will be held at the same venue with the same agenda of activities used for the pilot study, with the exception of adjustments made based on the pilot study results. This section describes the ALS procedures and tasks Pearson will implement during both the pilot study and operational ALS meeting and includes information about the configuration of panels and materials, training of panelists, the collection of panelists’ judgments, and the feedback given to panelists. The operational ALS will involve 33 panelists, 11 more than the number used for the field trial and pilot study. The TACSS will be consulted prior to the operational ALS meeting, and up to two members of the TACSS will be asked to observe the meeting. All operational ALS materials will be based on data from the 2017 grade 4 NAEP writing operational administration.

As previously indicated, Pearson will implement a standard setting procedure using the Moodle platform to facilitate key aspects of the BoW standard setting process including panelist training, review of ALDs, housing and review of student BoWs, recording of panelists’ judgments during each round of rating, provision of feedback, and evaluation of the standard setting process. The Moodle site will be set up to guide panelists through the steps of the standard setting process, with facilitators having the ability to restrict or provide panelists with access to sections of the site or activities, as needed. The night before the first day of the pilot study and the operational ALS meeting, Pearson will reset the Moodle passwords for all panelists to a meeting password. This password will be input by Pearson staff at the beginning of each day of the pilot study or operational ALS meeting to ensure panelists cannot access secure ALS materials from a tool (e.g., tablet, phone, or computer) other than the ones used for the meeting.
Preparing for the ALS Meeting

This section describes key activities Pearson will complete before the operational ALS meeting that contribute to the success of the meeting.

**Use of NAEP-Like Scales.** Pearson will create a NAEP-like scale(s) to avoid the risk of a premature release of achievement level cut scores for grade 4 NAEP writing. As in past ALS studies, the NAEP-like scale will be a linear transformation of the NAEP reporting scale. The transformation will be reviewed and evaluated by TACSS and the COR prior to implementation.

**Calculation of Student-Level Scores for each Body of Work.** Given that student-level results are not reported for NAEP, Pearson will place student work on a score scale for use in the BoW standard setting method. Pearson will compute an EAP ability estimate for each student using the same computational procedures implemented in 2011 for grades 8 and 12 NAEP writing achievement levels-setting (Measured Progress, 2012b). This approach will be evaluated by and discussed with the TACSS prior to implementation.

**Division of Panelists into Subgroups.** The NAEP ALS process has been implemented using a split panel design since 1991. For all three components of the grade 4 NAEP writing ALS process (field trial, pilot study, and operation ALS) the overall panel will be divided into two panels, referred to as subgroups, within the set of panelists selected. Pearson will assign panelists to subgroups with the intent of creating groups that are as equivalent as possible to one another. The subgroups have two general purposes: (1) to reduce the overall burden of the rating process by assigning a subset of 30 BoWs to each subgroup and (2) to provide replication of the ALS procedures across two subgroups. The panelists in each subgroup will be assigned to three table groups (two for the pilot study) with five panelists per table group.

**Selection of Bodies of Work and Assignment to Panel Subgroups.** The process for selection and use of forms and for selection of BoWs within forms is similar to that used for the field trial. Two sets of BoWs, selected from student responses to the 2017 operational administration, will be used for
both the pilot study and operational ALS meeting. Selecting the two sets of BoWs (each comprised of two responses by the same student) needed will, again, involve first identifying a subset of forms from which to sample the BoWs. Each form of the grade four NAEP writing assessment is comprised of two writing tasks representing two of the three purposes identified for the NAEP writing tasks—to convey, to explain, and to persuade. For grade four, 35% of the tasks focus on ‘to explain,’ 35% on ‘to convey,’ and 30% on ‘to persuade.’ The tasks are distributed across a number of task types identified by the type of stimuli used for the task.

Pearson will select eleven forms to ensure that all 22 tasks are represented and will assign seven forms to each panelist subgroup, three forms will be common across the two subgroups and four will be unique to a subgroup. The common forms will include at least one of each type of prompt and are intended to contain the tasks identified for public release as well as the other tasks paired with those prompts on the original forms.\(^3\) Table 1 provides an example of the distribution of tasks across common and unique forms for the two panel subgroups using a coding scheme that assigns the grade 4 tasks a code by purpose and task number (to convey = C1-C8, to explain = E1-E8, and to persuade = P1-P6). Assuming one task is released for each writing purpose, let C1, E1, and P1 in Table 2 represent the released tasks.

---

\(^3\) Prompts preliminarily identified for release will be used for this process as final decisions regarding the prompts selected for public release will occur after selection of the forms and BoWs for the pilot. If the final set differs from those used to select BoWs for the pilot, adjustments to the samples will be made prior to the operational ALS.
Table 1. NAEP Grade 4: Form Assignment by Panel Subgroup

<table>
<thead>
<tr>
<th>Group A</th>
<th></th>
<th>Group B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Common</td>
<td>C1E2</td>
<td>Group B</td>
<td></td>
</tr>
<tr>
<td>Forms</td>
<td>E1P1</td>
<td></td>
<td>P4E3</td>
</tr>
<tr>
<td></td>
<td>P2C2</td>
<td></td>
<td>C8P6</td>
</tr>
<tr>
<td>Unique</td>
<td>P3E4</td>
<td></td>
<td>C4E7</td>
</tr>
<tr>
<td>Forms</td>
<td>C7P5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E5C5</td>
<td></td>
<td>E6C6</td>
</tr>
<tr>
<td></td>
<td>C3E8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The assignment of forms will ensure that each group is assigned the same number of forms and the same number of tasks for each writing purpose; and that both across all tasks and within tasks for a particular writing purpose the average difficulty (i.e., mean score) of the tasks is approximately equal across the two groups. Pearson might need to make adjustments to the assignment of tasks and forms based on whether the paired prompts create a form administered in 2017 and meet the criteria for balancing difficulty across the two groups.

The first set of BoWs will then be selected from the forms identified for use and will serve as the BoWs for the first two rounds of the pilot study and operational ALS meeting, mirroring the process used for grades 8 and 12 (Measured Progress, 2012b). A total of 78 BoWs will be used for Rounds 1 and 2. Pearson will select the BoWs from the subset of student responses for which the student received a total raw score of 2-12, with each response scored 1-6, and neither response coded ‘non-scorable,’ ‘off-task,’ or ‘blank.’ We will, again here, use a stratified sampling process to select seven BoWs from each form with the exception of the common form that contains two released prompts, eight will be selected from this form. BoWs will be selected to ensure a uniform distribution across the scale of EAP ability estimates calculated for individual BoWs, that there is not a concentration of BoBs from a given form in any score range, and, to the extent possible, that there is representation of forms, tasks, and levels across the scale. Panelists in each subgroup will review 50 bodies of work, 22 of which will be common across both subgroups. Pearson content specialists, prior to the pilot study, will identify and remove student responses that could be distracting or that could
interfere with the rating process due to, for example, sensitivity issues. The content experts will also identify and remove BoWs for which the two student responses are disparate in terms of level and quality (e.g., one is rated at the high end of the scale and the other at the low end of the scale).

An independent set of BoWs will be selected from the same set of forms, for use in Round 3 of the pilot study and operational ALS meeting. The 78 sets of student responses in each set of BoWs will be divided so each panel subgroup will review 50 sets of student responses with 22 sets overlapping across the two panels (selected from the common forms) and 28 sets unique to each panel (selected from the independent forms). The selection process will mirror that used for the first sample with the exception that the BoWs used in the first set will be excluded from the sampling process for the second set of BoWs. The inclusion of independent sets of student responses for Round 3 provides an additional source of reliability data through the replication of the ALS process to a new and comparable set of BoWs. As with the field trial, for all rounds of the ALS process, the BoWs will be presented in descending order according to student performance based on the 2017 grade 4 NAEP operational assessment.

Provision of Advanced Briefing Materials. As discussed in Section 3 Briefing Materials, panelists will have access to advanced materials through the Moodle platform. Pearson will send each confirmed panelist access to the materials for their review and familiarization prior to the operational ALS, including information to log into the Moodle site and change the assigned password. The first time the panelist logs in to the website, he or she must read and electronically sign a nondisclosure agreement. Once signed, the panelist will be guided through a brief online training for using the Moodle platform before having access to the non-secure advanced materials and will use an online checklist to progress through the review and ensure each document is reviewed.

Training of Facilitators. The ALS process for the NAEP grade 4 writing assessment will be led by both a content facilitator and a process facilitator. The content facilitator was selected for her
expertise and experience in writing instruction. The process facilitator was selected for her expertise and experience conducting ALS meetings.

Recognizing that facilitators may introduce individual differences that can result in slightly different instructions, both the content and process facilitator will be trained to implement the process uniformly and as intended. PowerPoint presentations will be developed and reviewed by the COR and the TACSS. Facilitators will use these presentations to guide the ALS meetings. In addition, facilitator handbooks will include examples of the tables and graphs to be presented to panelists, a script for providing instructions, a description of each ALS activity, and an explanation of the feedback provided to panelists.

Facilitators will participate in the planning and development of the pilot study and the operational ALS meeting, and will attend a three-hour training prior to both of these studies focused on ensuring consistent understanding and implementation of planned processes. The project director overseeing the ALS activities will lead the training. In addition, the facilitators and project director will do a walkthrough of the entire ALS meeting the day before the operational ALS meeting. Panelist and observer feedback will be evaluated throughout the ALS panel meetings to inform facilitator processes and presentations. Adjustments will be made, as needed, to ensure the quality and effectiveness of the guidance and support facilitators will provide to panelists.

Preparing Panelists to make ALS Judgments

Provide an ALS Orientation. The operational ALS meeting will begin with the introduction of the project staff and a description of the panelist recruitment process. This will be followed by a presentation by the Governing Board Assistant Director for Psychometrics, providing history of and context for NAEP, NAEP ALS procedures, and NAEP writing assessments. The Governing Board Assistant Director for Psychometrics will also share the achievement level results from the 2011 grades 8 and 12 NAEP writing assessment (e.g., percent of students at or above each achievement
level) and from the most recent administration of the grade 4 NAEP writing assessment in 2002 (under the previous framework).

**Overview the Achievement Levels-setting Process.** Next, facilitators will provide an overview of the purpose of ALS in general and a description of the process used in setting the achievement standards.

**Review the Assessment.** Panelists will take the NAEP grade 4 writing assessment under conditions similar to those of the actual grade 4 NAEP writing administration. This activity will use the same customized software on the same Microsoft Surface Pro 3 or 4 tablets with Windows 10 operating system and detachable keyboards used by students taking the 2017 grade 4 NAEP writing assessment. Panelists will thus experience the assessment as the students experienced it and become familiar with the content and rigor of the assessment.

After completing the assessment, panelists will be trained in how to use the scoring rubrics developed to score student responses. Following training on the rubrics, panelists will receive scoring guides for the particular form used in their assessment and given time to review their own responses relative to those guides. They will be instructed that the purpose is not to score their work but to understand how responses were scored. This will be followed by a group discussion of the assessment experience and application of the scoring rubrics.

**Review the Framework and ALDs for Writing.** Facilitators will provide panelists with training in the key components of the NAEP Writing Framework and the rationale supporting these components. This training segment will include providing panelists with an understanding of the grade 4 NAEP writing ALDs, grounding them in what students should know and be able to do for the grade 4 NAEP writing assessment, and will conclude with a brief overview of the meeting schedule of activities.

**Orientation to Moodle Use.** The remainder of the computer-based ALS steps will take place on laptops provided by Pearson and will use the Moodle online software. Facilitators will guide panelists
through an orientation to show them how to use the online system for each step in the ALS process. Panelists will be trained in how to access student work samples through the Moodle site by selecting links to individual student samples and how to enter their judgments into the Moodle system. Pearson will incorporate questions into the Moodle interface to ensure the panelists’ readiness for completing a judgment and to confirm that panelists’ judgments are complete and entered as intended before submission.

After submission of panelists’ practice judgments, Pearson will review and analyze these judgments to provide panelists with appropriate feedback. Panelists will receive feedback through an online folder system within Moodle. Throughout the standard setting process, panelists will complete process evaluations to provide feedback to facilitators and to document the effectiveness of the process and the panelists’ preparation to complete the different aspects of the ALS procedures.

**Review Achievement Level Descriptions (ALDs).** As part of their online briefing materials, panelists should have reviewed the NAEP Writing Framework, the NAEP policy ALDs, and the NAEP grade 4 writing ALDs before attending the meeting. Facilitators will review these materials thoroughly during the operational ALS meeting to ensure panelists clearly understand the ALDs and to achieve general agreement across all panelists regarding the meaning of the ALDs. Panelists will be asked to think about the criteria for performance in each level described by the ALDs and then to focus in on the skills that differentiate student performance between levels. Panelists will have the opportunity to evaluate a set of student responses (independent of those used for setting achievement levels) and to place each response in an ALD category for discussion. In addition, facilitators will guide panelists through a review of the ALDs for grades 8 and 12 to understand the progression of skills needed as students advance through the grade levels.

**Collecting Content-Based Judgments**

During the writing assessment, students respond to two writing prompts. As previously noted, a student response to the two prompts will be considered the ‘Body of work’ that the ALS panelists will
be evaluating. BoWs will be housed in the secure Moodle platform. Panelists will review each BoW and classify the set of responses as below Basic, Basic, Proficient, or Advanced. These classifications will be used to calculate a recommended cut score from each panelist.

Pearson will follow the 2011 ALS implementation of the BoW method used for grades 8 and 12 NAEP writing (Measured Progress, 2012a), which excludes pinpointing in Round 3. In place of the pinpointing procedure, Round 3 will have panelists rate a new set of student responses, adding a replication component to the ALS process and providing further evidence for the evaluation of validity of ALS outcomes. Logistic regression will be used to calculate cut scores based on panelists’ ratings of each body of work into an achievement level (below Basic, Basic, Proficient, or Advanced). When used in standard setting, logistic regression establishes a logistic curve by setting up dichotomies representing whether a work sample is classified below or above each achievement level. This curve then represents relationship among the EAP scores and panelist’s ratings for all BoWs evaluated. An estimate of the panelist’s cut score is determined using the inflection point on the logistic curve. All feedback from this process will be provided to panelists using the Moodle platform.

Pearson will implement a data verification process on a sample of the information transferred at the following points in the ALS process:

- Panelists transfer of ALS judgments from paper forms to the Moodle interface.
- System transfer of ratings and responses from Moodle to the tools used for analysis.
- System transfer of feedback from the tools used for analysis to Moodle.

**Practice Round.** Facilitators will guide panelists through a practice round of making judgments about students’ BoWs. Panelists will independently review six BoWs and make a judgment about the achievement level classification of each. They will then discuss them as a table group in regard to the skills required in the ALDs and represented in the BoWs. The facilitator will then lead a whole group discussion on how to use the ALDs to classify BoWs into achievement categories.
**Round 1 Ratings.** Prior to beginning Round 1, panelists will complete a readiness survey through which panelists indicate their understanding of and comfort with the task. Facilitators will review the survey responses prior to beginning Round 1 and consult with any panelists who show a lack of understanding of or comfort with the identified task. Each panelist will then review 50 BoWs, covering the possible score range of the grade 4 NAEP writing assessment. The BoWs will be ordered from the highest to lowest performance, based on EAP score estimates. BoWs will be presented, reviewed and classified using the Moodle platform. After completing Round 1, each panelist will be asked to complete a survey addressing comfort with and understanding of the BoW standard setting process implemented.

Logistic regression will be used to calculate cuts scores based on panelist’s classification of each BoW into an achievement level (below Basic, Basic, Proficient, or Advanced). One cut score will be calculated for each achievement level for each panelist, thus each panelist will receive three cut scores after each round of ratings. The EAP student score computed for each BoW will be used as the independent variable in the regression analysis. The three dependent variables represent classification into Basic or above, Proficient or above, or Advanced or above.

**Round 1 Feedback.** After all panelists have completed their judgments for Round 1, Pearson will analyze panelist input and provide feedback using the Moodle platform (see Section 7 for examples of Round 1 feedback). Panelists will receive:

- Individual cut score information—each panelist’s cut score and where it falls in the distribution of cut scores,
- Rating group cut score information—the median, mean, minimum, and maximum cut score recommendation of each rating group, and
- Panel level cut score information—the median, mean, minimum and maximum cut score recommendation.

After reviewing the feedback data, panelists will be asked to discuss the bodies of work that
resulted in the most disagreement in judgments. Pearson will determine which bodies of work will be discussed after a review of the Round 1 results. Sets of student responses that require further discussion include those that:

- Appeared in the most achievement levels
- Had the largest spread of achievement level classifications
- Had a split such that panelists classified approximately the same number of Bows into two or more achievement levels
- Had a reversal (for example, lower scoring bodies of work were classified into higher achievement levels than higher scoring bodies of work)

**Round 2 Ratings.** Prior to beginning Round 2, panelists will complete a readiness survey through which panelists indicate their understanding of and comfort with the task. As with Round 1, panelists will be supported if they indicate a lack of understanding of or comfort with the process. Each panelist will then use the feedback from Round 1 as well as the discussions of the BoWs that generated the most disagreement, to provide judgments again for the same set of BoWs used for Round 1. Panelists will perform these tasks using the Moodle platform. After completing Round 2, each panelist will be asked to complete a survey addressing comfort with and understanding of the BoW standard setting process implemented and the information they relied on most to make their Round 2 judgments.

Again, logistic regression will be used to calculate cuts scores based on panelist’s classification of each BoW into an achievement level (below Basic, Basic, Proficient, or Advanced).

**Round 2 Feedback.** Panelists will receive the same types of individual and panel level feedback provided after Round 1, based on Round 2 data. In addition, they will receive consequences (i.e., impact) data based on the panels’ Round 2 judgments. These data will show the percentage of students at or above each median panel cut score, using the results of the 2017 grade 4 NAEP writing assessment. The consequences data will be presented using the Moodle platform in a way that will allow panelists to see how the percentage of students at or above a given cut score would change if the panelists classified
Facilitators will ask panelists to consider whether the grade 4 data seem reasonable given the knowledge and skills described in the ALDs and what panelists know about the distribution of knowledge and skills described in the NAEP writing framework. Facilitators will remind panelists to (1) take a national rather than a local perspective as the consequences data are based on a national sample and (2) to use the grade 4 writing ALDs, along with the feedback and discussions incorporated into the process, as the primary bounds for their Round 3 determinations.

**Round 3 Ratings.** Prior to beginning Round 3, panelists will complete a readiness survey through which panelists indicate their understanding of and comfort with the task. As with previous rounds, panelists will be supported if they indicate a lack of understanding of or comfort with the process. Each panelist will then use the feedback from Round 2, as well as their discussions with other panelists, to provide judgments for a new, but comparable, set of BoWs. Panelists will perform these tasks using the Moodle platform. After completing Round 3, each panelist will be asked to complete a survey addressing comfort with and understanding of the BoW standard setting process implemented and the information they relied on most to make their Round 3 judgments.

**Consequences Questionnaire.** The median panel cut scores from Round 3 will be the final cut score recommendations from the ALS panel. Panelists will be provided with consequences data based on their ratings from Round 3. In addition, panelists will receive feedback showing the distribution of whole panel cut scores and a list of the student BoWs reviewed during the meeting classified into particular achievement levels, using the final cut score recommendations to define the achievement levels. Panelist will review the consequences data using the Moodle platform, which will include a tool allowing panelists to manipulate the consequences data to see how that data, including the classification of BoWs, would change if the cut scores based on panelists judgments of the BoWs reviewed in Round 3, had been different than those resulting from the Round 3 determinations. After considering the consequences data through this process, panelists will complete a consequences questionnaire, asking
them if they recommend changes to one or more of the cut scores resulting from Round 3. For each recommended change, a panelist will be asked to provide a rationale for the change that is based in the BoWs reviewed, the ALDs, and the feedback and discussions they experienced as part of the ALS process.

At each step in the ALS process, panelists can view their entries in Moodle and will be asked to check them for accuracy prior to submitting their information. Once information is received through Moodle, data quality checks will be conducted to evaluate if the panelists’ input is logical. Unexpected or questionable input will be checked with the panelists before data summaries are computed and feedback is presented. All analysis programs will be developed prior to the operational ALS meeting, and two psychometricians will independently verify the program code. After running programs during the meeting, results will be verified through replication.

Section 7: Information Processing and Feedback

Pearson understands the importance of providing useful feedback to the many different groups involved in the ALS process. Section 7 describes the use of technology (hardware and software) to support efficiency and to strengthen the validity argument for the ALS outcomes; provides samples of the feedback described in Section 6; and describes the analysis plan for data obtained from all ALS studies as well as the planned strategy for examining and displaying results for presentation to the COSDAM and the full Governing Board.

Computer Use and Software

To support an efficient process and to strengthen the validity argument for the ALS outcomes, Pearson will use two computers during the field trial, pilot study and the operational ALS meeting. The use of computers and an online software interface will reduce the time required for panelists to complete most steps in the ALS activities. In addition, the use of computers will allow the panelists to interact with the items as grade 4 students did, such that panelists are better able to understand what examinees would have to know or be able to do.
Using the first computer, a form of the NAEP grade 4 writing test will be administered so that panelists can take the assessment using the same NAEP tablets that were used for the 2017 operational administration of the assessment. The assessment administration contractor is responsible for the configuration of the tablets. Pearson is responsible for the setup and administration of the tablets during the ALS meetings. Pearson will work closely with the assessment administration contractor to ensure that each panelist’s experience is as close as possible to what a student experienced during the actual administration.

The second computer will be a laptop computer (provided by Pearson) that is used by panelists to access the online Moodle interface. As a part of supporting both efficiency and accuracy of the ALS process, Pearson will use computers with access to the Pearson Moodle Standard Setting website, with a section specifically designed for the grade 4 NAEP writing ALS meeting. The grade 4 NAEP Moodle website can be accessed only through a user identification code (ID) assigned by Pearson. Permissions will be set up for each user ID so each panelist can access only the materials he/she will be using. The use of the Moodle site will reduce the time required for panelists to complete multiple steps in the process, since they will be able to access materials, including student responses, and make their judgments within the Moodle interface. As a website-based system, all materials and data will be stored, organized, and accessed through the Moodle site, which will ease the demands of development of materials and ensure the consistency of the materials with which the panelists interact. The use of the Moodle system will also ensure the security of the materials during the standard setting meeting, since the system will be used throughout the ALS process and requires a secure login by all users. Pearson designed the Moodle system for the NAEP grade 4 writing ALS process to have the following features:

- Simultaneous access by multiple users, with each individual user assigned a profile which defines their level of access to the site, including the materials the user can access and the specific functionalities available. Facilitators and site administrators can use conditional
access features based on date, activity completion, or custom settings to define access to materials and functionalities of the site.

- User access prior to the standard setting meeting, so users can interact with materials and experience training that will prepare them for the meeting and the ALS process.
- Consistent user interface throughout the entire ALS process, from pre-meeting work to post-meeting feedback.
- Management of storage and access to materials through the Moodle site, as a single access point for the meeting.
- Assignment of panelists to materials and sets of student responses, based on, based on panelists grouping. Panelist access student responses within the site by selecting a link within the site, easing their ability to move between student responses.
- Embedded quality control features through conditional settings within the item judgment survey that limit the range of responses to only valid entries and ensure no blank entries are accepted.
- Export of panelists’ item review judgments after each standard setting round is complete, as a group or individually. Facilitators will be able to view the panelists’ judgments within the Moodle site, to verify panelists’ responses.
- Secure storage of all panelist information and judgments within the Moodle site, with access restricted to facilitators and site administrators.
- Import and display of the results of response analyses, such as descriptive statistics for each panel group and for the panel as a whole, so panelists can view results of each round.
  - Panelists will access reports through folders on the site, which will only be accessible after each round has been completed.
Panelists at a table will have access to a panel level report containing only the information for the particular panel group to which the table belongs. On the panel group level report, the cut score recommendation for each panelist will be associated with the panelist’s unique user ID.

Panelists will have access to an interactive bar chart, representing consequence or impact data, after Round 3 of the ALS process. They will be able to adjust the recommended cut scores and see the consequences that would result from the changes imposed. That is, as a panelist moves the recommended cut score for an achievement level, the bar chart will adjust and update the percentage of students that would be in each achievement level.

**Sample Feedback**

Moodle can display various types of feedback data to panelists between rounds of ratings (see Section 6 for additional detail on the ALS process). Feedback will be provided to panelists at the completion of each round of ratings. This feedback will include individual cut score information, rating group cut score information (the median, mean, minimum, and maximum cut score recommendation of each rating group), panel level cut score information (median, mean, minimum and maximum cut score recommendation), and consequences data (or impact data) showing the percent of students at or above each achievement level. The following figures provide examples of the types of feedback that will be used.
Cut Score Location Data

Individual- and Group-Level Feedback

NAEP Consequences Data
Strategies for Results Presentation

In considering the recommended cut scores from the achievement level standard setting process, Governing Board members are likely to be most concerned about the reasonableness of the recommended cut scores and consequences data and the whether the achievement level process provided panelists with what was needed for them to make solid, content-based recommendations informed by grade 4 NAEP writing ALDs and supported by NAEP consequences data. The strategies described below will provide the information necessary for COSDAM and the full Governing Board to evaluate the results of the ALS process in light of these key considerations.

Strategy for Examining Results with COSDAM. As the technical subgroup of the Governing Board, COSDAM includes some members with significant psychometric and technical expertise who Pearson anticipates will be interested in developing detailed and well-supported explanations, on behalf of the full board, in regard to the two key considerations outlined above. Thus, Pearson recommends the following sequence for the COSDAM presentation:

1. Provide a high level overview of the BoW process implemented.
2. Present final recommended cut scores with the grade 4 NAEP writing consequences data.
3. Provide summaries of the feedback data for each round of the operational ALS process.
4. Provide context for these recommendations using summary information from other sources (e.g., ACT Aspire, SBAC, PARCC).
5. Provide summaries of panelist responses to all questionnaires, as part of considering procedural validity.

Strategy for Examining Results with the Full Governing Board. Pearson developed the strategy outlined below based on the assumption that the Governing Board, comprised primarily of policy makers, will be interested in a well-reasoned but nontechnical argument addressing the two key considerations described above. Thus, Pearson recommends the following sequence of steps for the full Board presentation:
1. Provide a high level overview of the BoW process implemented.

2. Present final recommended cut scores with the grade 4 NAEP writing consequences data.

3. Provide context for these recommendations using summary information from other sources (e.g., ACT Aspire, SBAC, PARCC).

Information presented to COSDAM can be used to respond to questions and concerns expressed by Board members.

Section 8: Identifying Potential Exemplar Student Responses

Identification of potential exemplar student responses is an important component of the ALS process as these student responses provide examples of the knowledge and skills required to perform at each level of achievement for NAEP writing and support the interpretation and use of NAEP results. Potential exemplar responses will be identified from the sets of student responses to the prompts on the common form for which both prompts are identified for public release. Facilitators will instruct panelists to decide whether student responses aligned with a particular achievement level “should be used,” “might be used,” or “should not be used” as an exemplar of performance within that achievement level. Pearson, in consultation with the Governing Board, NCES, and the TACSS, will further specify the process for securing panelist recommendations of exemplar responses that are consistent with the plans for reporting the 2017 NAEP writing results.

Section 9: Public Comment

Pearson understands the potential contribution of public comments to the development of and validity regarding the ALS procedures implemented and to the interpretation and use of the outcomes of that implementation. Pearson has designed a strategy for collecting and using public comment for this Design Document.

Pearson plans to implement a vigorous and targeted approach to soliciting public comment regarding the Design Document through personal contact with key leaders and members of stake
holder groups, and through the use of a simplified and directed format for reviewer response.
Pearson will create a website to obtain public comment on the design of the grade 4 writing ALS process. The website will provide a means for stakeholders and the public to find information about the Design Document and to provide feedback. Prior to opening the site to the public, Pearson will submit the site to Governing Board staff for review and approval.

Pearson will solicit comments from content persons, technical persons, and persons in education-related organizations that are known to use NAEP ALS data. When organizations are involved, a personal call from project staff will establish the initial contact and explain the importance of the request. For collection of comments from members of organizations, Pearson will ask the organization leadership to communicate the request for public comment to the membership and to encourage members’ cooperation and response. In all cases, a personalized email message will be sent to explain the purpose of the request for comment and to provide a link to the website having the Design Document, instructions, and questions for respondents to consider. Reviewer comments can be recorded through the link, downloaded by Pearson staff, and saved for analysis. Follow-up emails will be sent to encourage responses, and to contact individuals for clarification and additional information, should that be necessary. The public comments will be summarized and presented to the TACSS for discussion. TACSS recommendations will be shared with the Governing Board staff and the Design Document will be revised accordingly.

**Public Comment from Content Persons.** Pearson will contact various content organizations and ask for their feedback on specific aspects of the Design Document. For example, given that the National Writing Project has been especially helpful in previous NAEP writing achievement levels-setting projects, Pearson will ask for their assistance to identify at least one individual in each state, from among the 200 university sites, for review and feedback regarding specific aspects of the Design Document such as the use of a computer based assessment of writing for fourth grade students and the timing and content of the advanced materials. We will ask the leadership to provide feedback and to
send a message to encourage responses from content persons they identify. While these questions
clearly have a technical component, these are all questions that persons experienced in writing
instruction will have the necessary expertise to address.

In addition, Pearson will contact individuals who served on the 2011 Writing ALS panel for
grades 8 and 12 to ask for their comments. The process to be implemented for grade 4 is based on the
process implemented for grades 8 and 12 and while it will have been over five years since these
panelists were engaged in the NAEP ALS process, we believe they can provide useful information
regarding specific aspects of the current design.

**Public Comment from Measurement/Technical Persons.** Pearson’s staff has strong ties and
relationships with the measurement community and will contact members of the Board of Directors and
Chairs of Committees for the National Council on Measurement in Education (NCME) and leadership
persons for Division D of the American Education Research Association (AERA), for example, to ask
for their comments on specific aspects of the Design Document.

We also will contact persons who have made presentations about standard setting at NCME and
AERA during the past five to 10 years as well as the assessment director in each state to encourage his
or her participation. These individuals will be asked to comment on any aspects of the Design
Document that are of interest to them, and we will specifically request that they comment on the choice
of standard setting methodology (Body of Work) and the design of the pilot and operational studies.
While these areas are relatively broad in scope, this set of reviewers has the knowledge and expertise to
address these design elements.

**Public Comment from Organizations that Use NAEP ALS Data.** Several organizations are
very familiar with the work of the National Assessment Governing Board and with the achievement
levels for NAEP. Those organizations regularly use NAEP data in their reports, and Pearson will ask
the leadership of these organizations to provide comment on specific aspects of the design. While
reviewers in this group will be invited to comment on any aspect of the Design Document, Pearson will target the specific aspects that align with the organizations primary use of NAEP data. The following organizations are representative of those Pearson plans to include in this effort: The Education Trust, The Council of Great City Schools, The Fordham Foundation, The Lumina Foundation, the Council of Chief State School Officers, Achieve, American Federation of Teachers, National Education Association, The American Enterprise Institute, and The Gallup Institute. We will work with the staff of the Governing Board to finalize the list of organizations contacted.

**Section 10: Validation Studies**

Pearson’s approach to collecting validity evidence will focus on the three general categories of validity evidence that are typically recommended for standard setting: procedural, internal, and external (Sireci, Hauger, Wells, Shea, and Zenisky, 2009; Cizek, Bunch, & Koons, 2004; Hambleton, 2001; Kane, 1994, 2001). Procedural evidence refers to the appropriateness of the standard-setting procedures and the fidelity with which those procedures were implemented. Internal evidence refers to the internal consistency of data generated within the standard-setting meeting. External evidence refers to the relationships between interpretations based on the ALS outcomes in comparison to interpretations based on outcomes from a different, but directly relevant, outcome.

Pearson will describe plans for collecting and using each of these three types of validity evidence.

**Examination of Procedural Validity**

Procedural validity evidence refers to the appropriateness of the standard-setting procedures and how well those procedures were implemented. Evidence for procedural validity may come from a number of sources, including criteria for selecting panelists, the justification for the ALS method, the quality of the implementation of the procedure, and the completeness of the documentation of the process (Sireci, Hauger, Wells, Shea, and Zenisky, 2009). Pearson will provide all of these sources of evidence for evaluating procedural validity. In addition, panelists will be asked to complete evaluation forms after each round of standard setting and each major activity of the ALS process. Evaluations will
include both selected-response and open-ended questions that address the panelists understanding of the process and confidence in the results.

No key-entry by staff is required because panelists will use the secure online Moodle interface to complete their evaluations. Panelist entries will be available for viewing using the facilitator login to Moodle. Facilitators will review information from the panelists’ evaluations in real time throughout the ALS process, identify areas where panelists indicate a lack of understanding and/or comfort with the process, and address identified areas with specific panelists or the group as a whole.

**Examination of Internal Validity**

Internal validity evidence refers to the internal consistency of data generated within the standard-setting meeting. Pearson will conduct extensive statistical analyses on the cut scores to assess the internal consistency of the medians for grade 4 NAEP writing and the overall quality of the ALS process. The statistical analyses will be conducted using the NAEP-like scale scores. The same set of analyses will be completed for both the pilot study and the operational ALS meeting. Additional analyses will be completed comparing results for the pilot study and operational ALS meeting. Specific analyses will be used to examine the central tendency of the cut scores, change in cut scores by round, mean absolute difference, standard error of the cut scores, in addition to comparison of mean and median cut scores. Pearson will analyze cut scores for differences across panelist variables such as gender and race/ethnicity and process variable such as rating group.

**Central Tendency of Cut Scores.** Pearson will compute and report both mean and median panel cut scores for each round for rating groups. However, the median panel cut score will ultimately be used to calculate cut scores for each round of the ALS process. Medians are used rather than means because the median is less sensitive to outliers than the mean. Outliers may occur for a number of reasons, including the inexperience of panelists or panelists trying to intentionally influence the panel cut scores. Use of the median will help prevent these influences. However, mean cut scores will also be provided to panelists for complete information and in comparison to the medians. The percent of
students at or above each achievement level will be computed using the 2017 NAEP Grade 4 Writing Assessment operational results and the median cut scores for each round.

**Change in Cut Scores.** Pearson will report the change in cut scores by round. This provides additional information about how panelists were responding to the feedback provided. Pearson will compute the number and percent of panelists whose cut scores increased, decreased, or had no change from the previous round.

**Measure of Mean Absolute Difference.** Pearson will compute the mean absolute deviation as a measure of the variability of judgments within a panel. This measure will be as the average absolute difference between each panelist’s cut score and the median cut score.

**Standard Error of the Cut Score.** The standard error of a cut score is an estimate of the uncertainty in the reported cut score due to various sources of error. The reliability of cut scores emerging from a standard-setting process is typically regarded in terms of how consistent the cut scores are across panel/rater groups, and panelist type. This consistency across replications is used as an estimate of reliability. Due to the difficulty of calculating the standard error of a median, Pearson will report two estimates of the standard error of the cut scores that have been used in past NAEP ALS studies: an empirical-based method and a bootstrap method.

The first estimator of the standard error is based on consistency of cut scores between two rating groups using the same ALS procedure (Brennan, 2002). The second estimator of the standard deviation of the median is based on the bootstrap technique (Efron & Gong, 1983). In this procedure, repeated samples with replacement are taken from the original distribution of cut scores, and the median is calculated for each resample. The standard deviation of these medians is then calculated and used as the estimate. Pearson proposes to sample 1,000 cases to implement the bootstrap method.

Pearson recommends that the reported standard errors for rounds two and three of standard setting be interpreted with caution. Panelist judgments for these rounds are no longer independent due
to the standard setting process. Panelist judgments for Rounds 2 and 3 are influenced by the location of the cut scores for the other panelists and other factors that are part of the standard-setting process.

**Analysis of Means.** No satisfactory method exists for estimating the significance of the differences between groups on their median cut scores. However, if the mean and median cut scores are similar, then Pearson recommends that analysis of the effects on means be performed using an ANOVA procedure. Pearson will analyze cut scores for differences across panelist variables such as gender and race/ethnicity, and process variables such as panel/rating group.

**Examination of External Validity**

External evidence refers to the relationships between interpretations based on the ALS outcomes in comparison to interpretations based on outcomes from a different, but comparable, outcome. In exploring potential data sources to present to the Governing Board along with the ALS results, it was important to consider sources which tested similar populations with comparable tasks. Pearson has investigated three grade 4 writing/literacy assessments as potential data sources for an external validation check of the cut scores that will be established during the Body of Work ALS process for grade 4 NAEP writing. The ACT Aspire writing assessment and the PARCC and SBAC literacy assessments have been considered based on communicative purposes measured, task content and demands, response time allowed, rubric score scale, and the criteria/methods used to evaluate student responses.

With regard to communicative purposes, NAEP aligns better with Aspire than it does with PARCC or SBAC. In NAEP, three purposes—to persuade, to explain, and to convey experience—are measured. While Aspire at grade 4 measures only analytical expository writing, the “explain” task on NAEP is similar to the expository task on Aspire in that no reading is required prior to writing. Prompts are written to elicit an explanatory response that reflects students’ own thinking about the specified topic. That is, students must be able to clearly explain what they think about the topic, but
they are not required to bring particular background knowledge or use specific sources in developing their essay.

Whereas NAEP and Aspire are measures of writing only, PARCC and SBAC tasks are measures of literacy because they combine the assessment of reading and writing. Reading is required prior to writing the essay, and prompts are written to elicit an explanatory response that is based upon what students have read. Students must be able to clearly explain what they think about the topic, and they must use multiple sources from their reading to develop their essay.

There are also dissimilarities between the administration time of tasks. While on NAEP, students are administered two writing tasks, only one writing task is administered on Aspire. However, both NAEP and Aspire give students 30 minutes to respond to a writing task. In contrast, PARCC gives students 90 minutes to respond to the literacy task, which includes time to read, respond to selected-response questions, and write; and SBAC gives students 105 minutes, 35 minutes to complete the reading section and 70 minutes to complete the writing section, which is based upon what students have read.

The rubric score scale used for NAEP is similar to Aspire, though Aspire has no explicit counterpart to score point 6 in NAEP. A 5-point rubric is used for grades 3–5 to allow only one degree of performance above “adequate.” The rationale Aspire uses for this rubric scale is that finer distinctions above “adequate” are not evident in student responses in the lower grades. A 4-point score scale is used for both PARCC and SBAC.

In addition, the criteria used to measure NAEP and Aspire are similar in that NAEP assesses development, organization, language facility, and conventions; and Aspire assesses ideas and analysis, development and support, organization, and language conventions. However, NAEP uses a focused holistic scoring process, making a judgment about the overall response based on a consideration of the criteria, while Aspire uses an analytic scoring process, assigning a specific score to each criterion and adding these scores to derive a scale score and overall achievement level.
Despite the fact that there is evidence of some areas of misalignment between NAEP and the other three measures, based on COSDAM recommendations, Pearson will examine all three sources as comparative measures for external validity. This information will be shared only with COSDAM, the Governing Board, and Governing Board staff associated with this project.

Brennan, R.L. (2002, October). *Estimated standard error of a mean when there are only two observations.* (CASMA Tech. Note No. 1). Iowa City, IA: University of Iowa, Center for Advanced Studies in Measurement and Assessment.


