

National Assessment Governing Board

Developing Achievement Levels on the National Assessment of Educational Progress for Writing Grades 8 and 12 in 2011 and Grade 4 in 2013



Submitted to:
Dr. Susan Loomis
National Assessment Governing Board
800 North Capitol Street, NW, Suite 825
Washington, DC 20002-4233
Email: Susan.Loomis@ed.gov
Phone: 202.357.6940

This study was funded by the
National Assessment Governing Board
under Contract ED-NAG-10-C-0003.

Submitted by:
Measured Progress
100 Education Way
Dover, NH 03820
Phone: 603.749.9102

Table of Contents

Introduction	5
Purpose of This Document	5
Organization of This Document	6
Methodology	7
Body of Work.....	7
BoW Technological Integration and Enhancements (BoWTIE)	10
Research Studies	13
Field Trial	13
Hardware	14
Room Configuration	15
Test Administration	17
Static Information.....	17
Presentation of Student Work Samples	18
Field Trial Methodology	19
Field Trial Panelist Recruitment and Selection Process	22
Evaluation	28
Pilot Study	29
Special Study	29

The 2007 Writing NAEP	30
Review of Prompts and Scoring Rubrics	31
Classification of Student Work	31
Analyses	32
Achievement Levels-Setting Process	34
Selection and Recruitment of Panelists	34
Staged Selection Process.....	36
Use of BoWTIE in the Selection and Recruitment Process	50
Incentive Program	51
Assigning Panelists to Rating Groups.....	51
Assigning Forms to Table Groups and Selection of Student Work Samples ..	52
Assigning Forms to Rating Groups	54
Selection of Student Work Samples	56
Advanced Materials	57
The ALS Meeting	59
Facilitation	64
Room Setup.....	65
Panelist Training.....	67
Rounds of Rating and Feedback Information	71

Consequences Data Questionnaire.....	81
Selection of Exemplar Items.....	81
Evaluation	82
Security Procedures for Meetings.....	83
Public Comment Forums	84
Public Comment Initiative #1	85
Public Comment Initiatives #2 and #3	87
Reporting of Results	89
Validity Evidence	92
Procedural Validity.....	93
Internal Validity	94
Decision Accuracy and Consistency	95
References	97
Appendices	99

Introduction

On September 23, 2010, the National Assessment Governing Board (Governing Board) awarded a contract to Measured Progress for developing achievement levels on the National Assessment of Educational Progress (NAEP) for writing grades 8 and 12 in 2011 and grade 4 in 2013. This Design Document outlines the procedures for setting the cut scores corresponding to the three NAEP achievement levels (Basic, Proficient, and Advanced) and selecting exemplar writing prompts and student responses. Working with WestEd as a subcontractor, Measured Progress will conduct achievement levels-setting (ALS) meetings and produce a set of recommendations for the Governing Board to consider when establishing achievement levels on the NAEP for writing grades 8 and 12 in 2011 and grade 4 in 2013. Additionally, all aspects of the ALS process will be established with guidance from the Contracting Officer's Representative (COR), advice from the Technical Advisory Committee on Standard Setting (TACSS; see Appendix A for the list of members), and input from relevant stakeholders.

Purpose of This Document

The purpose of this Design Document is to describe the procedures for assisting the Governing Board with the development of achievement levels on the NAEP for writing grades 8 and 12 in 2011 and grade 4 in 2013. The Design Document describes the process to be implemented and the types of staff required to do the work. A customized Body of Work (BoW) standard-setting method will be used as the ALS methodology. Our goal is to implement a computer-based standard-setting method to enhance efficiency and effectiveness of the standard-setting procedure.

Organization of This Document

The Design Document is organized into six major sections. The **Methodology** section describes the methodology for conducting the standard setting and includes two subsections: the Body of Work (BoW) method and BoW Technological Integration and Enhancements (BoWTIE). **Research Studies** includes descriptions of the field trial, the pilot study, and the special study. The **Achievement Levels-Setting Process** section describes all aspects of the procedures, including nominating and selecting panelists, assigning forms to rating pools, distributing advanced materials, and a full description of the ALS meeting. The **Public Comment Forums** section describes the process designed by WestEd for three initiatives to gain input from a wide range of constituents regarding the design for setting achievement levels and the outcomes of the ALS processes. **Reporting of Results** describes the process of documentation and final reporting to provide access to NAEP achievement results to both the Governing Board and the broad audience of constituents for NAEP. Lastly, the **Validity Evidence** section discusses procedural validity and internal evidence to support inferences made based on the results of the ALS process. Additionally, decision accuracy and consistency based on cutpoints resulting from the process are discussed.

As prime contractor, Measured Progress will oversee all project planning, communications, schedules, budgets, and all subcontractor activities. As a subcontractor, WestEd will conduct public comment forums and will be responsible for all activities related to publicizing NAEP writing ALS plans and collecting input from NAEP stakeholders. WestEd will also handle logistical arrangements for pilot study and ALS meetings. Measured Progress will select and recruit panelists, make their travel

arrangements, and prepare panelists for their rating tasks. All reporting deliverables will be the responsibility of Measured Progress with input from WestEd as appropriate.

Methodology

Body of Work

Measured Progress will implement the BoW method for the NAEP writing ALS process. The BoW method belongs to the holistic family of standard-setting methods in which the panelist rating task consists of reviewing a series of examinee work samples and assigning each sample to one of several performance categories (Hambleton & Pitoniak, 2006). The BoW method (Kingston, Kahl, Sweeney, & Bay, 2001) is the method deemed most appropriate for writing assessments, as it was developed specifically for use with performance assessments that are designed to measure student achievement using open-response items.

The BoW standard-setting process includes an orientation and introduction to the assessment along with the purpose of the standard-setting meeting, a detailed review of the BoW method, training in the BoW classification tasks, and three rounds of rating examinee work samples, each followed by a process evaluation and presentation of feedback based on the rating round. Each of these components is described in detail in the Achievement Levels-Setting Process section of this document. Here, a description of the rating tasks is presented.

The rating tasks for the BoW method involve two distinct phases: rangefinding and pinpointing. In the rangefinding phase, student booklets covering the entire range of possible scores are presented for classification. Based on the cutpoints resulting from the rangefinding phase, the pinpointing phase uses only work samples in the vicinity of

the rangefinding cutpoints to focus more precisely on the performance that best represents the standard. The first two rounds of rating involve rangefinding while the third round focuses on pinpointing.

In round 1, panelists recommend cutpoints individually without discussion. In round 2, they recommend cutpoints individually, following extensive group discussion. In round 1, panelists work individually with the Achievement Level Descriptions (ALDs) and the set of work samples ordered from lowest to highest expected a posteriori (EAP) score. NAEP writing examinee work samples consist of the responses to the two writing prompts administered to the student. For each work sample, the panelists note the knowledge, skills, and abilities (KSAs) demonstrated by the student. This step enables panelists to familiarize themselves with work samples across the full range of possible scores. Once the panelists are finished working their way through the work samples individually, without consulting with their colleagues, they are asked to decide which achievement level (Basic, Proficient, or Advanced) the work sample best represents. Work samples that cannot be classified into one of the achievement levels will be considered Below Basic. While the work samples are presented in order of EAP score, panelists are not required to rate them in strictly increasing order. Instead, panelists are encouraged to take a holistic look at the examinee work sample, rather than making a judgment based primarily on the ordering of the work samples.

After all panelists complete their ratings, individual cutpoints are calculated using logistic regression. The group's cutpoint is the median of individual panelists' cutpoints. The median is the central tendency statistic of choice for this purpose because it is less susceptible to the effects of extreme values.

In statistics, logistic regression is a model used for prediction of the probability of occurrence of an event by fitting data to a logistic curve. In standard setting, an event consists of a panelist's classification of a work sample. By setting up dichotomies, denoting whether a work sample is classified below or above each achievement level, a logistic curve can be established. This logistic curve represents the empirical relationship among the total scores of all work samples and a panelist's ratings. The inflection point of the logistic curve corresponds to an estimate of the panelist's cutpoint. For each panelist, a logistic curve is fit for each cutpoint and the estimate for each group's cutpoint is the median across panelists.

Once cut score feedback has been presented to panelists, group discussion begins by starting with the first work sample for which there was disagreement as to how it should be rated or classified; the panelists discuss the categorization of the work samples according to their round 1 ratings. Panelists are encouraged both to share their own point of view as well as to listen to the thoughts of their colleagues. The goal is to allow each panelist the opportunity to explain why he or she classified each work sample into one achievement level or another. Facilitators make sure the panelists understand that the purpose of the discussion is not to come to consensus, as at every point throughout the standard-setting process, panelists will be asked to provide their own individual best judgment. Panelists are also told which work samples the median cut scores fall between. Once the discussions are complete, the panelists complete the round 2 ratings using the same set of examinee work samples.

Once panelists complete their round 2 ratings, cutpoints are calculated in the same manner using logistic regression, and panelists are again given information on

which work samples the median cut scores fall between along with other feedback that is described later in this document. Before the final round of ratings, the set of examinee work samples are modified to remove those that are clearly within an achievement level and are replaced by additional samples with scores around the cutpoints set from the previous round. This pinpointing round allows panelists to narrow in on final cut score recommendations. With this set of work samples, panelists are asked to complete the same rating tasks as described for the first two rangefinding rounds. The benefit of the pinpointing round is that examinee work samples are selected based on the location of the round 2 cutpoints such that the final cut scores recommended contain a higher degree of precision. Logistic regression is once again employed to calculate final cut score recommendations.

To aid in the BoW standard-setting process, a computer-based tool has been designed specifically for the NAEP assessments. This tool will enhance the adequacy and efficiency of the BoW assessment method. The design of the tool is described in the following section.

BoW Technological Integration and Enhancements (BoWTIE)

The computer-based standard setting designed for 2011 and 2013 NAEP writing is ideally suited for the BoW method. Measured Progress will use BoWTIE, an integrated and technologically enhanced implementation of the BoW method. A fully computer-based system allows greater ease in developing and preparing materials, ensuring consistency of materials among panelists, and simplifying the organization of materials. The wholly computer-based standard setting is both cost-effective and environmentally sensitive as the need for hardcopy materials is minimized. Panelists

will access and annotate materials and enter their ratings in a database as a natural extension of computer-based assessment. The use of BoWTIE will also enhance security of the materials during standard setting.

BoWTIE will integrate the entire ALS process including: recruitment of panelists, selection of student work samples, dissemination of advanced materials and information, panelist training, rounds of rating, feedback, and process evaluations. An additional feature of BoWTIE is the capability of providing interactive consequences data feedback. The integration of all parts of the standard setting will enhance efficiency, security, and replicability of the ALS process.

More importantly, the specific tool built for this contract is designed to meet all the technical and statistical adequacy criteria set by Berk (1986). In fact, BoWTIE addresses a limitation originally cited: ease of implementation. Ease of implementation is no longer a limitation when using this computer-based tool, as it eliminates the inconvenience of preparing for the pinpointing round. When results from rangefinding are computed and approved by the COR to be used for the next round, BoWTIE selects student work samples in real time around the cutpoints resulting from rangefinding.

The BoWTIE designed for the NAEP assessments has the following features:

- multiple users with different
 - levels of access
 - types of information provided
 - functionalities
- fail-safe features for quality assurance
- interactive feedback

- manual override if there are outliers
- personal annotation features
- sorting capabilities
- consistent interface for panelists before, during, and after the ALS meeting
- database of panelist information
- storage of advanced materials
- instantaneous reports from process evaluation
- print features for accommodation

Another key advantage to a wholly computerized standard-setting process is the ability to allow panelists to focus more on their ratings and less on managing the vast amount of materials that are customarily distributed at a standard-setting meeting. Because panelists will be entering their ratings on the computer, data analysis occurs automatically after panelists finish their ratings. The built-in quality assurance features will also ensure that all ratings are within range and no blanks are left before panelists leave the rating session.

BoWTIE provides panelists annotation and navigation tools to enable them to view various response booklets without having to shuffle through large stacks of printed test booklets. Panelists can view actual student responses by simply clicking on a booklet number. This feature eases their ability to move from one student test booklet to the next and to flip back and forth so comparisons can be made between and among student responses.

Research Studies

In preparation for the ALS, three research studies will be conducted: a field trial, a pilot study, and a special study. The field trial intends to address the implementation of an entirely computer-based procedure and process, as this represents a significant change relative to previous ALS procedures. The pilot study is standard ALS process and is conducted to ensure the smooth implementation of the operational levels-setting methodology. Finally, the special study's purpose is to provide appropriate information to explore the relationship between performance on the 2011 assessment, based on the new writing framework, and performance on the 2007 assessment, based on the writing framework first implemented for the 1998 NAEP. A description of these research studies follow.

Field Trial

For the purpose of testing the logistics of using an entirely computer-based system, a small-scale field trial will be conducted by implementing selected parts of the process designed for the operational ALS meeting. Measured Progress will critically investigate all logistical aspects of the computer-based implementation during the field trial. The procedures implemented in the field trial are an abbreviated version of the procedures designed for setting the 2011 NAEP writing achievement levels. The field trial will use a simplified, scaled-down version of the ALS sampling process to select a single panel of 20 for the two-day study. It is our intention that the field trial will serve as a trial run for the full ALS procedures as they relate to the computer-based implementation of the BoW method. To this end, the field trial will be conducted at an offsite hotel venue, following a series of in-house trials at Measured Progress and user

acceptance testing of BoWTIE. The field trial will allow us to emulate the 2011 procedures, identify any logistical weaknesses, and adjust the procedures as necessary for further evaluation in the pilot study where the exact ALS operation procedures will be carried out. The field trial will focus on the logistical elements of the meeting that are directly impacted by the use of computers. In particular, our evaluation will center on five main elements: (1) hardware, (2) room configuration, (3) test administration, (4) presentation of static information, and (5) presentation of student work samples. Although each of the above logistical elements will be thoroughly tested prior to the field trial by an internal group at Measured Progress, the field trial will serve as an operational investigation of each element during an actual implementation of the ALS process.

Hardware

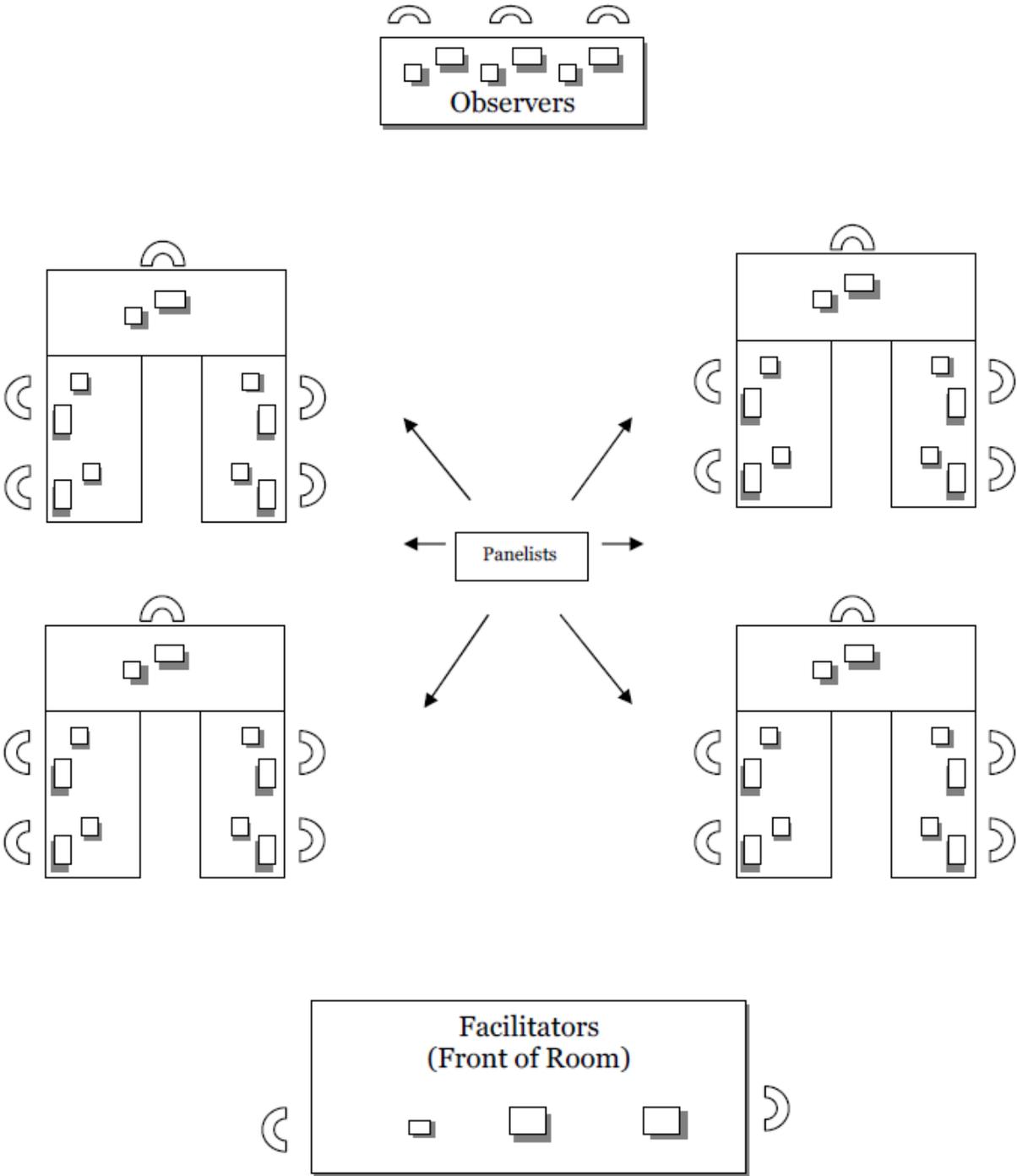
Measured Progress staff is responsible for the transportation, storage, and networking of the netbooks. This will include ensuring that appropriate equipment (extension cords, Ethernet cords, routers) is available and working properly onsite. The Office of Information Technology (OIT) Infrastructure staff from Measured Progress will configure the netbooks to be hardwired to a local server. Particular attention will be given to the security of the netbooks. Specifically, the netbooks will be configured to allow access only to the standard-setting server. This will limit distractions from emails, internet, etc., and eliminate security breaches. Each of the above elements will be carefully scrutinized by the appropriate Measured Progress staff (OIT and Program Management) to ensure optimal configuration for the achievement level setting.

The test administration contractor is responsible for delivering the NAEP laptops, setting them up on-site, and dealing with any issues that arise with the NAEP equipment during the standard-setting process. Measured Progress will work closely with the test administration contractor to make certain the panelists' experience of taking the test on the NAEP laptops is as close as possible to the student test-taking experience.

Room Configuration

A critical aspect of any ALS effort is the room configuration for the meeting. The room must be set up to facilitate discussion while simultaneously allowing the panelists to work independently. For this purpose, the room will be configured with four sets of three six-foot tables arranged in an open square to seat five panelists per table with all panelists able to face the front of the room. Site visits were conducted to examine a mock setup of the rooms to verify that these specifications can be used. Figure 1 displays the room configuration diagram sent in advance of the site visit. In addition to the configuration of the tables, particular attention will be given to the placement of all cords and cables. The safety of the room will be of paramount importance. An extra day will be scheduled at the beginning of the meeting to set up, configure, and test the computers and other equipment in advance. The number of staff and amount of time required to set up, test, and pack up the equipment will be carefully evaluated in order to optimize efficiency.

Figure 1. Room Configuration for Field Trial for NAEP Writing



Test Administration

It is critical to the success of standard setting to give panelists the opportunity to fully comprehend the test instrument and its administration to students. To achieve this goal, each panelist will be using a NAEP laptop computer to take a form of the NAEP writing assessment. Measured Progress staff will observe a test administration of NAEP writing, request administration directions, and attend a special session where Westat will demonstrate the computer distribution, setup, and administration. These activities will be conducted to help ensure that panelists receive the same spoken instructions and tutorials as the students who took the NAEP. The goal is to make sure the panelists' experiences taking the NAEP replicate the students' experiences.

An aspect of the ALS procedures that might present a challenge is the logistics involved in having each panelist use two computers. Panelists will be using the same laptop computer to view all of the writing prompts in the test forms assigned to them. For all other aspects of the process, panelists will be using netbooks configured optimally for BoWTIE. During training, particular attention will be given to ensure smooth transitions between computers as needed. The training process is designed to give panelists adequate information, experience, and time to familiarize themselves with the use of the NAEP laptops and the netbooks. Most importantly, the field trial will test the logistics of requiring panelists to manipulate two computers.

Static Information

The implementation of a computer-based process offers the distinct advantage of a paperless meeting. Before transitioning to a paperless environment, the ability of panelists to use the computer-based materials will be carefully scrutinized. To this end,

the field trial will be used as an opportunity to survey panelists to learn their preference for the way static information is displayed. Information distributed to panelists is considered static if it does not change throughout the standard-setting process. More specifically, it is information that is independent of the cut scores: ALDs, p-value feedback, and Reckase Charts. For the field trial, ALDs will be projected onto an overhead screen, accessible on individual computer screens, displayed on large posters around the room, and printed for each panelist. Reckase Charts and p-value feedback will be provided on the computer and printed for each panelist. Panelists will complete an evaluation form with questions regarding the ease and preference for the display of static information. Based on these evaluations, modifications may be made for the pilot study.

Presentation of Student Work Samples

The BoW procedure includes both a rangefinding and pinpointing stage. For the rangefinding stage, a set of pre-identified work samples is selected to represent the full range of achievement and a balance of passage type across forms. For the pinpointing stage, a second set of work samples is selected to target a finer range of the achievement continuum, based on the resulting cutpoints from the rangefinding stage, while still maintaining passage type balance across forms. BoWTIE aids in the student work sample presentation for the following ALS procedures:

- storing and presenting the rangefinding work samples
- ordering the work samples by their EAP score estimate
- selecting the second set of work samples targeted to the cutpoints identified during rangefinding

Panelists will be asked to evaluate the cognitive load and logistical ease involved in reviewing both rangefinding and pinpointing booklets using BoWTIE, as it is possible that the classification task will become more difficult for panelists when asked to make distinctions between work samples that represent similar achievement, especially in the pinpointing stage. Additionally, particular attention will be given to panelists' evaluation of the number of booklets assigned for review. Because of the expectations for longer responses written by 12th graders, the single-panel field trial will be implemented for grade 12 NAEP writing.

Field Trial Methodology

A single-panel standard-setting process will be conducted using either the data from the pilot administration of the 2011 assessment or electronic images of the 2007 NAEP student responses. The procedures employed will be an abbreviated version of those proposed for the pilot study and operational meetings. The agenda for the field trial is in Figure 2.

Figure 2. Agenda for 2011 NAEP Writing ALS Field Trial

Agenda	
2011 NAEP Writing for Grades 8 and 12	
Achievement Levels-Setting (ALS) Field Trial	
July 9–10, 2011	
Portsmouth, NH	

Day 1 – July 9, 2011	
8:00 AM	Registration
8:30	Welcome and Introductions
8:40	General Orientation to the NAEP
9:00	NAEP Exam, Scoring, and Discussion
10:30	Break
10:45	NAEP Writing Framework Review NAEP Writing ALDs Review
12:00 PM	Lunch
12:30	Round 1 Ratings (Range-finding) Evaluation #1
5:00	Adjourn

Day 2 – July 10, 2011

9:00 AM	Recap of Previous Day's Activities
9:15	Feedback from Round 1 Cut Scores Rater Location Feedback P-Value Feedback (Overall Difficulty of Each Prompt)
10:00	Break
10:15	Feedback (cont'd) Reckase Charts Consequences Data Evaluation #2
11:30	Lunch
12:30 PM	Round 2 Ratings (Pinpointing)
4:00	Evaluation #3 Debriefing
4:30	Adjourn

Excluding time for setup, the implementation of the field trial will take two full days. In comparison, the ALS meetings will span approximately three and a half days. Because the field trial is meant to test the logistics and computer-based components of the meeting and only grade 12 standard setting will be conducted, some agenda items have been abbreviated: introductions, framework and ALD reviews, training, and rating rounds. Thus, panelists will complete one round of rangefinding during the first day

followed by an evaluation (Evaluation #1) to assess both the ease of transitioning between computers and the logistics of entering ratings into BoWTIE. The second round of rangefinding will not be conducted. Instead, all cut score feedback will be presented on the second day followed by an evaluation (Evaluation #2) to assess which modes (on-screen, projected, displayed on posters, or printed) for presentation were most helpful. Panelists continue by providing their ratings for the pinpointing stage (round 2). This stage is followed by the final evaluation (Evaluation #3), which intends to assess the cognitive load and ease of the rating task during the pinpointing stage. Finally, panelists will be dismissed after a short debriefing of the two-day field trial.

Site for the Field Trial

The original location selected for the field trial was Measured Progress, Dover, NH. However, during the first TACSS meeting on December 2–3, 2011, a recommendation was made to change the location for the field trial to an off-site location that will require implementation of procedures with conditions more similar to those likely to be encountered in the operational implementations of the process. The logistics being examined include transporting, setting up, and packing up the equipment, which includes two computers to be used by each panelist. The field trial is scheduled for July 9–10, 2011, in Portsmouth, NH.

Field Trial Panelist Recruitment and Selection Process

Twenty panelists who possess appropriate content knowledge in writing for students in grade 12 will be recruited from southern New Hampshire and Maine for the field trial. Given the need for only two panels and a total of twenty panelists, all selected from a limited geographic area for the field trial, the demographic representation of

panelists for the field trial will be based on only three variables: panelist type (teacher, nonteacher educator, and members of the general public), type of educational institution (public school districts and private schools), and the qualifications of panelists.

Consistent with the intended panel composition for the ALS process, the selected panel for the field trial will be composed of 55% teachers, 15% nonteacher educators, and 30% general public members. It is intended that all panelists will be knowledgeable in writing and student performance in writing in grade 12. Teacher panelists will be classroom teachers in secondary schools who teach writing, composition, or journalism (hereafter termed writing) depending on the courses taught at their specific institutions. Nonteacher educator panelists may be curriculum specialists in writing or other educators with a background in writing; these educators are currently not classroom teachers in grades K–12. Furthermore, faculty members in writing at public and private two-year and four-year postsecondary schools will be considered as nonteacher educator panelists. General public panelists are members of the general public who are in a position in their professional practice to evaluate writing samples such as reports and general memoranda. Specifically, they are not current or former educators. They may be professional writers such as journalists, editors, or publishers.

The recruitment and selection of panelists involves the identification of districts, the identification of panelist nominators, contacting nominators, and, lastly, the selection and recruitment of panelists. These stages are described below.

Identification of Districts and Panelist Nominators

A database of school districts in New Hampshire and Maine will be used to identify 224 nominators within a 50-mile radius of the field trial site, Portsmouth, NH. The goal of 224 nominators assumes 30% of the nominators will respond to the request for nominations by submitting at least one nominee for consideration. Based on this assumption, 224 nominators should yield at least 67 nominees. Furthermore, assuming that 30% of the nominees will meet the panelist qualifications, accept the nomination, and be able to participate, 67 nominees will yield the final target of 20 panelists for the field trial. Potential nominators will be identified to nominate the type of qualified panelists in the proportions indicated in Table 1 below. Table 2 presents the targeted number of districts, institutions, and nominators as well as the number of panelists each nominator source is expected to yield.

Table 1. Demographic Classification for Panelists for the NAEP ALS Field Trial

Demographic Classification	Percentages
Panelist Type	55% Teachers 15% Nonteacher Educators 30% General Public
Private/Public	90% Public Schools 10% Private Schools
Qualification Score	Based on desired credentials for each type of panelists

Table 2. Number and Percentage of Nominators from Different Sources

		Nominators			
		Source	Number of Districts/ Institutions	Expected Number of Nominators	Percentage
Panelist Type	Teacher	Public School Districts	33	99	55
		Private Schools	24	24	
	Nonteacher Educators	Public School Districts	7	21	15
Private Schools		6	6		
Postsecondary Institutions		7	7		
General Public	Public School Districts	67	67	30	
		<i>Total Public School Districts</i>	<i>107</i>	<i>187</i>	<i>83</i>
		<i>Total Private Schools and Other Institutions</i>	<i>37</i>	<i>37</i>	<i>17</i>
		Total	144	224	100

To meet the goal of selecting 55% of the panelists who are teachers of writing, 99 potential nominators from public school districts and 24 from private schools will be identified within a 50-mile radius of Portsmouth, NH. The 99 potential nominators identified from public school districts will be district superintendents, school principals, school board presidents, heads of teacher organizations, or presidents of parent teacher organizations. The principals of the 24 private schools will be identified as potential nominators. The total of 123 nominators who will be asked to identify writing teachers from public and private school settings represents 55% of the total number of 224 nominators. If possible, each of the two panels will have a representative from a private school. Based on information from previous ALS processes, recruiting panelists from private institutions is historically quite difficult. Therefore, an 80:20 ratio, public school district to private school, will be used for selecting nominators in order to achieve our goal of a 90:10 ratio for panelists (again,

public school district to private school). A sample email to be sent to nominators of teachers appears in Appendix B.

To meet the target of 15% nonteacher educators for the field trial panel, district superintendents, school principals, and school board presidents will be identified as nominators from public school districts. School principals will be identified as nominators from private schools. In addition, chairs of the appropriate academic departments, such as interdisciplinary studies in writing or the school of journalism, and directors of writing centers or writing fellows programs will be identified as nominators from postsecondary institutions. The goal of this sampling effort is to identify 34 nominators of educators who are not teachers in K–12 classrooms and who have content expertise in writing at grade 12. These nominators will be asked to consider nominating themselves as well as other candidates who are knowledgeable about the current expectations for writing in high school. The letter in Appendix B will be modified appropriately to seek nominations for nonteacher educators.

To meet the goal of selecting 30% of the panelists from members of the general public, the education committee chairs of the Chamber of Commerce, mayors or town managers, editors-in-chief of the local newspapers, librarians, members of the school board and the human resource departments of large corporations, all located in Portsmouth and surrounding townships, will be identified as nominators. The appeal to these nominators will clarify the qualifications for nominees, including the knowledge of skills for writing general memos and evaluating writing samples. The letter in Appendix B will be modified appropriately to seek nominations for general public panelists.

Contacting Nominators and Selection of Panelists

The nominators described in the preceding sections will be asked via email to nominate up to four panelists with the qualifications detailed in Table 3 below. Nominees will be selected as panelists based on their self-reported qualifications in response to an online questionnaire. A five-point scoring scheme will be used to choose the most qualified panelists who are knowledgeable about writing while maintaining the goal to recruit 10–11 (55%) teachers, 3–4 (15%) nonteacher educators, and 6 (30%) members of the general public to compose the panel of 20.

Table 3. Qualifications of Nominees by Panelist Type

Teachers	Nonteacher Educators	General Public
At least five years teaching experience	Nonteacher educational staff at secondary schools, e.g. principals and writing coaches	Familiarity with writing at the relevant grade level
and	or	and
At least two years teaching writing, composition, or journalism courses	Curriculum directors and content specialists at secondary schools who have a background in writing, composition, or journalism or curriculum specialists at the state’s Department of Education	Not former employees of an educational institution; professional activities include evaluating writing samples
and	or	and
Exceptional performance as judged by a supervisor	Post-secondary faculty who teach writing, composition, or journalism	Engagement in professional activities involving writing

At the field trial, each panelist will be assigned to a table of five panelists with representation among panelist types at each table. In addition, an attempt will be made to assemble tables with equal numbers of male and female panelists and to balance other demographic variables according to the overall sample distribution.

The field trial sampling process described above will serve as a trial run for the sampling procedure planned for the pilot study and operational ALS meetings. In particular, the use of a questionnaire, the qualification scoring scheme, and some elements of the nomination and recruitment process will be tested prior to implementing the sampling process for the pilot study and ALS meetings. Any district selected for the field trial will not be disqualified from the nomination process for the pilot study and ALS meetings; however, individual panelists participating in the field trial who are subsequently nominated as panelists for the pilot study or ALS meeting will not be selected as panelists for these later meetings. Each field trial participant will receive a 10.1-inch netbook and will be reimbursed for travel expenses consistent with federal travel requirements.

Evaluation

The logistical implementation of the computer-based procedures will be evaluated by gathering panelists' appraisals of the activities during the field trial. Each of these elements will be assessed through the administration of a questionnaire as described earlier in this section. Specifically, we will gather panelists' appraisals of the amount of workspace, ease of computer operation and manipulation, the room setup, the elements of the meeting that worked best, and the things they would change.

Results from the field trial will inform the training process for the pilot and operational ALS meetings.

Pilot Study

A pilot study using the exact procedures designed to set the achievement levels for the writing assessments will be implemented. Procedural results from the pilot study will provide information regarding operational aspects of the procedure, feedback presentation, and the amount of time and understanding necessary for a smooth implementation of the operational levels-setting methodology. Measured Progress will conduct a pilot study for each grade in advance of its corresponding ALS meeting. There will be 20 panelists per grade who will be selected using the same nomination and recruitment scheme used for selecting ALS panelists. The goal is to assure that every detail of the pilot study is as similar as possible to the planned procedures for the ALS meeting. A thorough discussion of the ALS meeting is in a later section.

Special Study

The Governing Board has requested that a special study be implemented for the purpose of providing information for exploring the relationship between performance on the 2011 assessment, based on the new writing framework, and performance on the 2007 assessment, based on the writing framework first implemented for the 1998 NAEP. The special study described here is based on the recommendations from the first meeting of the TACSS and is not the study that was originally requested. Measured Progress will implement a special study during the pilot study with the purpose of exploring the relationship between performances on the two assessments. In addition

to the change in achievement level descriptions and cut scores that will result, a number of elements have also changed and need to be carefully considered. These elements include the transition from paper-and-pencil to computer-based administration, the prompts, the scoring rubrics, and the student population. In an attempt to mitigate the influence of the above elements, Measured Progress will have the pilot study panelists engage in a separate rating round to categorize 2007 student responses using 2011 ALDs. Both of the pilot panels for the 2011 NAEP writing, one for grade 8 and one for grade 12, will perform an additional round of ratings on the 2007 student responses, following the pilot study. Thus, each panel will participate in four rounds of ratings, three for the pilot study and one for the special study.

Upon completion of the pilot study, which involves implementing the exact procedures of the ALS meeting, the panelists will again come together as a group to be given an overview of the purpose of the special study and the steps in the process. At this point, the panelists will have experienced extensive training on the BoW procedures. Therefore, the overview will emphasize the abbreviated version necessary for the special study; namely, taking the 2007 writing NAEP, becoming familiar with the 2007 prompts and rubrics, and classification of student work. It is important for panelists to take the 2007 writing NAEP to allow them to experience the difference between taking the paper-and-pencil test (2007) and the computer-based version (2011). No feedback will be provided to panelists following the classification task.

The 2007 Writing NAEP

The special study will involve enough forms to ensure that all genres are represented in proportion to the framework specifications. All panelists will review the

same test forms and the same work samples. Panelists will be given a brief orientation to the test and the test-taking situation. Since the 2007 writing NAEP was not administered by computer, panelists will take the paper form of the test and review scanned images of handwritten student responses.

Review of Prompts and Scoring Rubrics

After panelists are introduced to the 2007 writing assessment, they will review all prompts and scoring rubrics for their grade level. Panelists will review KSAs required to respond to the 2007 prompts as they relate to the 2011 ALDs. To ensure common application of the 2011 ALDs to the 2007 prompts, panelists will be presented with a training set of student work samples at each score level for three selected prompts, one for each genre.

Classification of Student Work

Panelists will examine sets of work samples from the 2007 administration that are distributed across the score range and classify them into the Basic, Proficient, and Advanced levels based on the 2011 descriptions. The work samples will be scanned images presented on the computer, and the annotation functionality will be available. The panelists will input their ratings on the computer and will have an opportunity to review their ratings before submitting for analysis. Samples will be presented in rank order from lowest to highest score. No further training in the classification process will be provided, as this has been the panelists' task during the rating rounds of the pilot study.

Analyses

A series of cross-tabular analyses will be conducted in an attempt to understand the relationship between the performance on the 2007 assessment using the 2007 achievement levels and performance on the 2007 assessment using the 2011 ALDs. The goal is to compare the achievement level classification from the reported 2007 results to the achievement level classification using the 2011 ALDs. Because each student is not assigned an official achievement level classification, Measured Progress will run the comparison based on classifications from the student's plausible values as well as the classification that would have arisen had the student been assigned a single EAP score. In addition, the classifications based on the 2011 ALDs will be examined using the individual panelist classifications as well as the classifications that result when cut scores are calculated using logistic regression. The cross-tabular comparisons will be reported as follows:

Table 4A. Correspondence Between 2007 Achievement Levels Based on EAP and Panelists' Classifications Based on 2011 ALDs

Actual 2007 Classifications (EAP Estimates)	Special Study Panelists' Classifications (2011 ALDs)			
	Below Basic	Basic	Proficient	Advanced
Below Basic				
Basic				
Proficient				
Advanced				

Table 4B. Correspondence Between 2007 Achievement Levels Based on Plausible Values and Panelists' Classifications Based on 2011 ALDs

Actual 2007 Classifications (Plausible Values)	Special Study Panelists' Classifications (2011 ALDs)			
	Below Basic	Basic	Proficient	Advanced
Below Basic				
Basic				
Proficient				
Advanced				

Table 4C. Correspondence Between 2007 Achievement Levels Based on EAP Estimates and Panelists' Classifications Based on 2011 ALDs after Logistic Regression

Actual 2007 Classifications (EAP Estimates)	Special Study Panelists' Classifications (2011 ALDs) After Logistic Regression			
	Below Basic	Basic	Proficient	Advanced
Below Basic				
Basic				
Proficient				
Advanced				

Table 4D. Correspondence Between 2007 Achievement Levels Based on Plausible Values and Panelists' Classifications Based on 2011 ALDs after Logistic Regression

Actual 2007 Classifications (Plausible Values)	Special Study Panelists' Classifications (2011 ALDs) After Logistic Regression			
	Below Basic	Basic	Proficient	Advanced
Below Basic				
Basic				
Proficient				
Advanced				

The level of correspondence between actual classifications of student work samples and the classifications provided by the special study panelists will be the basis for judging whether performance between the 2007 and 2011 NAEP writing are

comparable. A high level of correspondence between the two classifications will be indicative of the comparability of performance between the two assessments and can inform later discussions of the reasonableness of the 2011 ALS results.

Achievement Levels-Setting Process

Measured Progress will conduct the ALS process by using the informed judgments of well-qualified and broadly representative panels to recommend achievement level cut scores consistent with the achievement levels definitions for Basic, Proficient, and Advanced levels and to identify exemplar performance for each level. The following describes how we will conduct the ALS process, including each step before, during, and after the ALS meeting.

Selection and Recruitment of Panelists

As panelist selection is a critical aspect of any standard-setting study (Cizek & Bunch, 2007; Raymond & Reid, 2001), panelists for the pilot study and ALS meetings will be nominated and selected according to a detailed plan designed to meet the Governing Board's specifications for a broadly representative panel. The most important panelist attribute to qualify for the task of setting achievement levels for NAEP in writing is knowledge in the subject area of writing and of student performance in the appropriate grade level. For the 2011 NAEP in writing, panels will be assembled to set achievement levels in writing for grades 8 and 12. For the 2013 NAEP in writing, panels will set achievement levels in writing for grade 4.

In addition to selecting panelists based on their knowledge in the subject area, the Governing Board specifies that each panel must be composed of teachers (55%), nonteacher educators (15%), and members of the general public (30%). In other words,

70% of the selected panelists will be educators and 30% non-educators. The panelist types were defined previously in the field trial section and their qualifications in Table 3.

One hundred panelists who possess appropriate content knowledge for writing for students in grade 8 and 12 will be recruited from across the nation to participate in the standard-setting process for the 2011 NAEP in writing. Forty (two grade-level panels of 20) of these panelists will be selected for the pilot study on November 15–18, 2011, in St. Louis, Missouri, and 60 (two grade-level panels of 30) for the operational ALS meeting on February 7–10, 2012, also in St. Louis. These panelists will be recruited simultaneously using a multi-staged sampling plan to assemble panels closely reflecting the demographics of the nation’s school population.

Fifty panelists will be recruited separately for the standard-setting process for the 2013 grade 4 NAEP in writing. Twenty of these panelists will be selected for the pilot study on November 19–22, 2013, and 30 for the ALS meeting on January 21–24, 2014. Each study will have a single panel for grade 4.

The description below is for the 2011 NAEP in writing for grades 8 and 12. A similar process will be implemented for the 2013 grade 4 NAEP in writing, although findings from the 2011 recruitment and selection process may suggest modifications in the process for the later ALS meeting.

A brief look ahead at the end goal for the composition of panels may be helpful before describing the steps in the sampling plan. For each grade-level pilot study, a panel of 20 will be formed with the following composition: 11 (55%) of the panelists will be teachers, three (15%) will be nonteacher educators and six (30%) will be

representatives from the general public. A grade-level panel of 30 members for the operational ALS meeting will be composed of 16 (55%) teachers, five (15%) nonteacher educators, and nine (30%) members of the general public.

Broad representation according to socioeconomic status (SES) and urbanicity of the student population and proportional representation by educators from private and public institutions will be addressed in the first stage of the selection process, which is the sampling of districts from which nominators will be identified. In addition, the demographic characteristics of the panelists will be proportionally representative of national population distributions based on gender, race/ethnicity, and geographic location. These attributes will be collected from nominated panelists in the last phase of the selection process, the selection of panelists. All stages of the recruitment and selection process are described in the following sections.

Staged Selection Process

Measured Progress will use a comprehensive multi-staged approach to select NAEP writing ALS panelists. Each stage of the selection process will ensure a broad level of representation:

Stage 1: Select districts and identify nominators

Stage 2: Contact nominators

Stage 3: Contact nominees

Stage 4: Select and recruit panelists

The goal of the selection process is to recruit panelists who will be able to make individual judgments and contribute collectively to the process of setting standards that

inform the nation about what students should know and be able to do at the assessed grade levels.

As listed above, the general sampling design for selecting panelists has several steps. First, a sample of districts will be selected. Nominators will be identified in the selected districts and invited by email to submit up to four nominations for panelists by completing nomination forms available electronically. Nominees will be notified of their nomination and asked to submit their credentials using an online form. Panelists will be selected giving top priority to the highest-qualified nominees. The use of specific databases, sampling variables, types of nominators, and panelist classification targets will be detailed below in sections describing the sampling for each standard-setting study.

Stage 1: Select Districts and Identify Nominators

Districts are the primary sampling unit for the sampling design. Using the 2008–2009 Common Core of Data (CCD), only districts with students at grades 8 and above (15,468 of the 18,350) will be considered (U.S. Department of Education (DOE), NCES, Institute of Education Sciences (IES), 2010a). The method of district selection is described next.

The goal of 100 panelists for the pilot study and ALS Meeting for the 2011 NAEP in writing requires identifying a much larger sample of nominators to account for those nominators who do not respond to the request for nominations and, subsequently, nominated panelists who either do not meet the desired qualifications or do not accept their nomination. Based on response rates reported in previous panelist recruitment efforts for the NAEP and, particularly the recent trend of decreasing response rates for

nominated teachers, the response rates used in this sampling plan will vary depending on the type of panelists being recruited.

For all panelist types, it is assumed that 30% of the nominators contacted will respond with at least one nomination; however, the conservative estimate of only one nominee is expected. A response rate of 12% for teacher nominees is estimated whereas a response rate of 30% is projected for all other nominees. Given that recruiting panelists from private institutions is historically quite difficult, an 80:20 ratio, public school district to private school, will be used for selecting nominators in order to achieve our goal of a 90:10 ratio (again, public school district to private school) for panelists. Based on these assumptions, 2,352 nominators will yield 110 panelists. This meets the goal of 100 panelists with adequate overage.

District sampling without replacement will be performed for the three panelist types: teacher, nonteacher educator, and general public. It is expected that each district in the teacher sample will yield three nominators; each district in the nonteacher sample also will yield three nominators; and each district in the general public sample will yield one nominator. In addition to district sampling, sampling of private schools and postsecondary institutions will also be performed as the first step before identifying the respective nominators. See Table 5 for a description of who may serve as nominators.

Table 5. Potential Nominators from Each Sampling Unit

Panelist Type	Sampling Unit	Nominators
Teacher	Public School District	Superintendent Principal School Board President Head of Teacher Organization President of Parent Teacher Organization (PTO)
	Private School	Principal
Nonteacher Educator	Public School District	Superintendent Principal School Board President State Curriculum Specialist
	Private School	Principal
	Postsecondary Institution	Chair of Appropriate Academic Departments (e.g., School of Journalism or Interdisciplinary Studies in Writing) Director of Writing Center or Writing Fellows Program
General Public	Public School District	Mayor City or Town Manager Education Committee Chair of the Chamber of Commerce Editor-in-chief of the local newspaper Librarian Members of the School Board Department of Human Resources and Directors for Corporations

Districts will be selected to proportionately represent the four NAEP regions, SES, and urban and rural demographics as presented in Table 6. The representation by NAEP region is based on the number of districts in each of the four regions. The

selected districts will include a representative sample of districts with low SES as indicated by the percentage of students who participate in the National School Lunch Program in the district (U.S. DOE, NCES, IES, 2010b; 2010c).

Table 6. Demographic Classification of Districts

Demographic Variable	Attributes	Percentages
NAEP region ¹	Midwest	35
	Northeast	20
	South	25
	West	20
SES ²	Low SES	20
	Not Low SES	80
Urbanicity ³	Large City	15
	Large Suburb	33
	Rural	20
	Other	32

¹ U.S. DOE, NCES, IES, 2011

² U.S. Census Bureau, 2011

³ U.S. DOE, NCES, IES, 2010b

The 2010 U.S. Census reports 20.7% of the U.S. children under the age of 18 lived at or below the poverty level in 2009 (U.S. Census Bureau, 2011). The top 20% of districts ranked from high to low based on the percentage of students enrolled in the National School Lunch Program will be taken as the target for districts with low SES.

The Common Core of Data (CCD) database identifies districts by size and urbanicity using the four categories of rural, town, suburb, and city, and classifying

each of these as small, medium, or large. The large city classification identifies 1,245 (7%) districts with a population of 250,000 or greater (U.S. DOE, NCES, IES, 2010b; 2010d). Fifteen percent of the students in grades 8–12 go to school in these districts. The Council of the Great City Schools grants membership eligibility to districts located in cities with populations over 250,000 (Council of the Great City Schools, 2011). The sampling of districts will target 15% of large city districts to assure their representation in the sample. In addition, 33% of the districts will be selected to represent the proportion of students educated in districts categorized as large suburbs, 20% to represent the percentage of students schooled in rural districts, and 38% to represent all other urban areas.

Identification of Nominators Using District Sample

To meet the goal of selecting 55% of the panelists who are teachers of writing, nominators will be identified from public school districts and private schools. Three or more nominators will be identified in each of 460 public school districts (selecting according to the process described in an earlier section) to yield at least 1380 nominators. These nominators for teachers will be district superintendents, school principals, school board presidents, or heads of teacher or parent organizations. In addition, 345 private schools will be selected randomly and the principal (titles may also be chancellor or headmaster) will be identified as a nominator. The ratio of nominators from public districts to nominators from private schools will be 80:20 while the respective desired representation of panelists is 90:10. This is due to the historically lower response rates from private schools. A 12% participation rate will be used for teacher nominees from public school districts. The final goal is to select 50

teacher panelists from public school districts and 5 teachers from private schools in keeping with the 90:10 ratio described above. A sample email to be sent to nominators of teachers appears in Appendix C. Table 7 presents the details of the expected number of teacher panelists from two sources: public school districts and private schools.

To meet the target of 15 nonteacher educators (15%), 212 nominators will be identified as follows: (1) three nominators (superintendent, principal, school board president, head of teacher organization, or president of PTO) in each of 34 public school districts to yield 102 nominators, (2) 20 nominators (principal, chancellor, or head of school) from each of 10 private schools, and (3) 100 nominators (e.g., department chairs) from each of 100 postsecondary institutions. The nominators from postsecondary institutions will be used primarily on grade-12 panels. These nominators will be encouraged to nominate themselves as well as other candidates who are knowledgeable about the current expectations for writing in grades 8–12. The letter in Appendix C will be modified appropriately to seek nominations for nonteacher educators. Table 7 presents the details of the expected number of panelists from the three sources described above.

To meet the goal of selecting 30% of the panelists who are not educators, 405 districts will be identified. From towns in the same locality, the chair of the education committee of the Chamber of Commerce, the mayor or city manager, the editor-in-chief of the local newspaper, a librarian, members of the school board, or the director of the human resources department or divisional director at large corporations will be identified as potential sources for nominators. The appeal to these nominators will clarify the qualifications for nominees including skills for writing general memos and

evaluating writing samples. The letter in Appendix C will be modified appropriately to seek nominations for general public panelists. Table 7 presents the details of the expected number of panelists to be recruited from the general public.

Table 7 provides a consolidated description of the number and percentages of nominators from different sources. Since the goal is to simultaneously select 100 panelists for the pilot study and ALS meeting for the 2011 NAEP in writing for grades 8 and 12, the panelists will be distributed appropriately between these two meetings. For example, the panelists recruited from postsecondary institutions primarily will be used on the panels for grade 12 for the pilot study and ALS meeting. In summary, a total of 2,352 nominators is expected to yield 707 nominees from which 117 panelists will be selected. The yield of 110 panelists provides a comfortable overage to ensure the availability of 100 qualified panelists for the two studies.

Table 7. Number and Percentage of Nominators from Different Sources

		Nominators					
		Source	Number of Districts/ Institutions	Expected Number of Nominators	Expected Number of Nominees	Expected Number of Panelists	Number of Panelists Needed
Panelist Type	Teacher	Public School Districts	460	1380	414	50	55
		Private Schools	345	345	104	12	
	Nonteacher Educators	Public School Districts	34	102	31	9	15
		Private Schools	20	20	6	1	
		Postsecondary Institutions	100	100	30	9	
	General Public	Public School Districts	405	405	122	36	30
	<i>Total Number Sampled from Public School Districts</i>		899	1887	567	95	
	<i>Total Number Sampled from Other Institutions</i>		465	465	140	22	
	Total		1364	2352	707	117	100

Identification of Nominators from Private School Sampling

To reach the goal of achieving proportional representation among educator panelists from private schools, a sample of 365 schools (345 for teacher nominators and 20 for nonteacher educator nominators) using random sampling without replacement will be drawn from the private schools listed in the NCES Private School Universe Survey (U.S. DOE, NCES, IES, 2010c) that offer grade 8 and higher (18,594 private schools offer grade 8 or higher of the total 28,450 private schools in the database). Additionally, private schools will be sampled to provide proportional representation across the four NAEP regions based on the number of private schools offering grade 8 and higher in each region. Table 8 presents the percentage of private schools in each NAEP region.

Table 8. Private Schools by NAEP Region

NAEP Region	% Private Schools
Northeast	24
South	31
West	25
Midwest	20

Identification of Nominators from Postsecondary Institution Sampling

One hundred colleges will be randomly selected from the College Navigator database published by NCES. The sample will include approximately 30 (30%) two-year and 70 (70%) four-year postsecondary institutions. The number of colleges to be selected will support the intent to increase the representation of postsecondary faculty

on ALS panels for grade 12. This is consistent with the efforts of the Governing Board to include “preparedness” in the NAEP assessments for grade 12.

Identification of Nominators from Other Sources

In a separate sampling, publishing companies and national professional associations of writers and journalists will be identified. In particular, the heads of publishing companies that publish young adult fiction will be asked to identify nominators. To ensure broad stakeholder participation in the process, the selection of nominators will include suggestions by officers of stakeholder groups such as the Association for Supervision and Curriculum Directors (ASCD), the Conference on College Composition and Communication (CCCC), the Conference on English Leadership (CEL), the Council of Chief State School Officers (CCSSO), the Council of Great City Schools (CGCS), the International Reading Association (IRA), the National Association of Elementary School Principals (NAESP), the National Council of Teachers of English (NCTE), the National Writing Project (NWP), and Teachers of English to Speakers of Other Languages (TESOL). Each stakeholder group will receive an invitation to nominate members of their organization who are residents of the selected districts.

Stage 2: Contact Nominators

Each nominator will be requested to nominate up to four persons of the appropriate type (teachers, non-teacher educators, or general public members) for each grade level. The qualifications for panelists will be specified in the email. These qualifications may be found in Table 3 (Field Trial section).

The email will direct nominators to a link where nominations can be submitted online. In the email, the nominator will be provided secure access to a website through a username and password. Once logged in, they will be asked to complete one online form for each nomination. On the form, they will be asked to provide the name and contact information of the nominee as well as a short description of the nominee's qualifications. Nominations will be immediately available in the BoWTIE database for use in the sampling process. If the nominator does not have an email address, they will be contacted by mail. They will be able to provide their nominations on paper forms.

Stage 3: Contact Nominees

Each nominee with complete information and proper qualifications will be entered in the nominee pool and will be contacted via email (or by letter if the nominee does not have an email address). In the email, the nominee will be provided introductory information regarding NAEP writing, NAGB, and the ALS process. They will also be informed about the incentive for participating in the ALS process, namely the receipt of a netbook computer. In the same email, they will be provided secure access to a website through a username and password. Once logged in, they will be asked to provide information relevant to their nomination and possible selection to be a panelist, including information regarding their qualifications. The website will include a functionality that will enable them to upload a copy of their curriculum vitae and links to additional information pertinent to the NAEP ALS process. Logging onto the website and providing requested information will signify their acceptance of the nomination and their elevation to candidate status.

Stage 4: Select and Recruit Panelists

Candidates will be selected from the nominee pool based on their self-reported qualifications and the needs for a balanced distribution of panelists. Results from the field trial may inform this panelist selection process. Please see Table 3 in the Field Trial section for the qualifications required of each panelist type. Each panel will be formed to maintain the stated proportions of educators and non-educators as indicated in Table 9. To the extent possible, panelists will be selected from the pool of nominees to approximate the proportional representation by other characteristics shown in the table, particularly by the geographical region, ethnicity, and gender of the panelists. Whenever feasible, panels are created to balance the remaining characteristics, namely by SES, urbanicity, and the number of panelists from private and public institutions.

A five-point scoring scheme will be used to choose the most qualified panelists from the pool of candidates. The scoring will be weighted to assure panelists who are knowledgeable about writing will receive the highest scores. Each panelist in the pool will be classified using eight variables. The variables are listed in Table 9 below.

Table 9. Classification of Candidates in the Nominee Pool

Demographic Variable	Attributes	Percentages
Panelist Type	Teachers	55
	Nonteacher Educators	15
	General Public	30
Gender	Female	50
	Male	50
Race/Ethnicity	Caucasian	80
	Non-Caucasian	20
NAEP region	Midwest	35
	Northeast	20
	South	25
	West	20
SES	Low SES	10
	Not Low SES	90
District Size	Large City	15
	Large Suburb	33
	Rural	20
	Other	32
Private/Public	Public Schools	90
	Private Schools	10
Qualification Score	1–5 based on desired credentials for each type of panelists	

Before the meetings, each panelist will be assigned to a table of five with balanced representation by panelist type, race/ethnicity, gender, and NAEP region. The ultimate goal is to create representative panels of educators and non-educators in whole, grade, rating, and table groups.

Use of BoWTIE in the Selection and Recruitment Process

The nomination and panelist selection process will be primarily web-based through BoWTIE. In all communications to nominators and nominees, it will be made clear that participants in this computer-based ALS process will be using netbooks with a 10.1-inch screen. As a backup method for nominators who either lack Internet access or simply prefer to work with paper and pencil rather than online, they will be allowed to submit nominations on paper. Nominees will be strongly encouraged to use the web-based tools, as studies have shown that use of these tools reduces data entry errors. This effort is consistent with Measured Progress's desire to maintain quality assurance in all data collection endeavors.

Once a person has been nominated, they are assigned a username and password and become an official user of BoWTIE. From that point on, each prospective panelist is in the nominee pool and, if selected, uses the same user information for the rest of the process (to receive advanced material, participate in training and rounds of ratings, and complete process evaluations, for example). During the nomination process, nominees will be directed to online surveys to provide key information including their education, educational experience, gender, race/ethnicity, geographic region, willingness to participate, and information regarding their qualifications.

Incentive Program

As described throughout this section, the design to use technology in all aspects of the ALS process will enhance efficiency and promote quality assurance. A significant element of this approach is for each panelist to use a netbook computer to receive training and information before rounds, input their ratings during the rounds, and receive feedback data after rating rounds. The netbooks selected for the ALS process have a 10.1-inch screen, 1 GB RAM, 160 GB hard disk drive, and a webcam. Each panelist for the pilot study and the ALS meetings will receive a netbook as an incentive. Both for procedural and security reasons, the netbooks will be returned to Measured Progress between meetings and will be reset with the manufacturer's settings. Due to these security measures, panelists will be unable to access information left by other panelists and NAEP materials and information will not be made public, whether intentionally or accidentally.

Assigning Panelists to Rating Groups

As required by Governing Board policy and described in the panelist recruitment process, 70% of the panelists for each grade will be educators and 30% will be non-educators. The educators will be further divided such that 55% of the panelists will be educators at the grade level for which standards are being set, and 15% will be nonteacher educators, such as postsecondary writing faculty teaching introductory courses or writing teacher education programs, and state and local curriculum directors. The groups shall reflect an overall balance of race/ethnicity, gender, and geographic location.

Each grade-level panel will be split equally into two groups, and each group will thoroughly review a subset of prompts with some prompts in common between the two groups of panelists. Effort will be made to maintain a similar distribution of gender, race/ethnicity, and geographic representation within each subgroup of panelists.

The panelists in each subgroup will be further assigned to a table group of five members allowing for more manageable discussions, an important part of the ALS process. Panelists will be assigned an identification number to identify them in BoWTIE. This identification number will also be used to track the secure materials that are distributed during the process. A second, secret identification number will be assigned to each panelist for the purpose of receiving specific feedback after a round of rating.

Assigning Forms to Table Groups and Selection of Student Work Samples

Selecting the student work samples that will be used for a BoW standard setting is in itself an important and time-consuming task. This task is further complicated by the test design of NAEP writing. Each student responds to two NAEP writing prompts requiring different communicative purposes. Based on the 2011 NAEP Writing Frameworks, the three communicative purposes of writing assessed for the 2011 and 2013 NAEP are to persuade, to explain, and to convey an experience. The distribution of writing tasks for each writing purpose for each grade is shown in Table 10. Table 11 reports the number of writing tasks for each writing purpose as well as for each type of prompt for grades 8 and 12.

Table 10. Percentage of Writing Tasks per Writing Purpose

Writing Purpose	Grade 4	Grade 8	Grade 12
To Persuade	30%	35%	40%
To Explain	35%	35%	40%
To Convey Experience	35%	30%	20%

Table 11. Number of Writing Tasks per Writing Purpose and Type of Prompt

Grade	Purpose for Writing	Total	Type of Prompt			
			Text	Visual	Audio	Video
8	Convey Experience	6	0	3	1	2
	Explain	8	1	5	0	2
	Persuade	8	0	5	0	3
12	Convey Experience	5	1	0	0	2
	Explain	9	2	4	0	2
	Persuade	8	0	5	0	4

Each of the two prompts in a booklet will be scored using a holistic scoring rubric with score levels ranging from 1 to 6. Since each student receives two prompts, the total raw score for each student booklet ranges from 2 to 12. Blank and off-task responses will not be included in the student work samples that will be rated. Furthermore, to enhance efficiency in the rating process, the TACSS recommendation was to avoid using student booklets with scores of (1, 1) and (6, 6) for the two prompts.

This is based on the expectation that these extreme scores, by default, are classified in their respective extreme categories.

There are two prompts in each test form, and 44 test forms for each grade. Each prompt occurs in exactly four forms, two times in each position. Note that with this configuration, only 11 forms are needed to exhaust all of the prompts. Thus, the intention is to use only 11 forms for the ALS process to minimize the physical demands on the panelists. Fatigue during the rating process can threaten the reliability of panelists' classifications and thus threaten the validity of the results. The test forms that will be used for the ALS process will be evenly distributed among the two panel groups according to the following criteria:

- Each group classifies the same number of test forms.
- The average difficulty of forms across groups is approximately equal.
- The number of each type of prompt is approximately equal.
- Three forms, selected to include at least one of each type of prompt, will be common across the two groups.
- Paired prompts correspond to test forms that were administered.

Assigning Forms to Rating Groups

There are 22 prompts in the 2011 NAEP writing test for each of grades 8 and 12. The number of prompts for each writing purpose (to convey = “C,” to explain = “E,” and to persuade = “P”) and the number for each prompt type (text, visual, audio, and video) are presented in Table 11. Consider the example of writing prompts for grade 8. Let the writing prompts be identified as C1 to C6, E1 to E8, and P1 to P8. One prompt for each writing purpose will be released to the public. Without loss of generality, let the

released prompts be C1, E1, and P1. Select C2 so that the average difficulty of C1 and C2 is in the middle of the difficulty range for all the “C” prompts. Now let E2 and P2 be selected similarly. The released prompts and the selected prompts will be included in the prompt rating pool for each rating group. The rest of the prompts will be assigned to the rating groups to maintain the same number of prompts for each writing purpose for each group and the same average difficulty of the prompts assigned to each group within a writing purpose. Using this prompt assignment process, the forms that will be rated by each rating group for grade 8 is presented in Table 12. A similar assignment scheme adjusted for the different numbers of prompts for each writing purpose for grade 12 is also in Table 12. Adjustments might be made to the resulting assignment of prompts and forms to rating groups depending on whether the prompts that are paired create a test form that was actually administered.

Note that four forms will be specific to each group and three will be in common. The total of 11 distinct forms will be used across groups such that seven forms are assigned to each group. Thus, each group is effectively assigned 14 prompts in the rating pool.

As recommended by the TACSS, 50 student work samples will be rated by panelists in each group. This will result in approximately seven selections from each form for the rangefinding stage of the rating process. For the pinpointing, where student work samples are selected around each cutpoint resulting from round 2 rangefinding, two or three student work samples will be selected from each form for each cutpoint for a total of 15 student work samples per cutpoint. Thus, a total of 45 student work samples will be rated by the panelists during round 3.

Table 12. Form Assignments to Rating Groups

Grade 8		Grade 12	
Group A	Group B	Group A	Group B
C1E2		C1E2	
E1P1		E1P1	
P2C2		P2C2	
C3E4	C4E3	C3P3	E3C4
P7C5	P8C6	E4C5	E5P4
E5P5	P4E7	P7E6	P6E7
P3E8	E5P6	E8P5	P8E9

The form assignment and sample selection will be finalized based on advice from the TACSS.

Selection of Student Work Samples

Student samples will be selected to be uniformly distributed across the score range. All prompts assigned to a rating group will be represented in the work samples that panelists will be rating. Each panelist will be asked to rate 50 student work samples in the rangefinding stage and 45 student work samples in the pinpointing stage. This number of student work samples each panelist rates in each round is deemed reasonable in the sense that panelists are not likely to become too fatigued to provide reliable and valid ratings. The plan for sampling student work and assigning it to the panelists is described next.

The final assignments of forms and booklets to each panelist for the rangefinding round will be stored in BoWTIE prior to the meeting. The booklets

selected for rangefinding and training will be reviewed by the content facilitators to make sure there are no student responses that could be distracting or interfere with panelists' ratings (e.g., sensitivity issues). The content facilitators will be trained in the use of BoWTIE and will be able to log in to the system to perform their assigned task of reading students' responses that are part of the training and rangefinding samples. Content facilitators will be trained in BoWTIE for this purpose subsequent to user acceptance testing and after sample student responses are received from the appropriate NAEP contractor through ETS, the design, analysis, and reporting contractor.

Pinpointing booklets will be selected onsite, based on cutpoints set by the panelists from their rangefinding round. All of the student responses and corresponding data received from ETS will be stored in the BoWTIE system for possible selection for the pinpointing round. Pinpointing booklets should not be duplicates of the rangefinding booklets. Once pinpointing booklets are selected, content facilitators will have the task of reading them before they are finalized for the next rating round.

Advanced Materials

Advanced materials are considered the first step of training panelists (Cizek & Bunch, 2007; Raymond & Reid, 2001). Raymond and Reid (2001) noted that training of panelists in NAEP ALS processes has been outstanding. Based on two decades of careful research conducted for NAEP ALS projects regarding the timing and amount of training materials, Measured Progress will uphold the tradition of past ALS processes. The timing and amount of materials may be adjusted based on findings from the pilot study.

Early formal notification of selection to panelists is important in making panelists comfortable about their roles in the activities. In addition to the customary letter of notification and/or cover letter, the materials they will receive are in Table 13. Appropriate materials will be available to panelists online four weeks prior to the meeting and in hard copy two weeks prior to the meeting, unless otherwise indicated. Note that some of these materials will be made available to them during the recruitment process for the purpose of providing information about the NAEP ALS project.

Table 13. Panelists’ Advanced Materials

Material	Online	Hard Copy	Comments
NAEP and/or the Governing Board Brochure	✓	✓	
2011 NAEP Writing Framework	✓	✓	
2011 NAEP Writing Specifications	✓	✓	
2011 NAEP Writing Final Achievement Levels Descriptions	✓	✓	
Confidentiality Agreement	✓		To be completed and signed during registration
Press Release Form	✓		Fill out online form
Briefing Booklet	✓	✓	Available online and hard copy simultaneously two weeks prior to meeting
Preliminary Agenda	✓	✓	Available online and hard copy simultaneously two weeks prior to meeting
Hotel Map and Directions	✓	✓	

The ALS Meeting

Due to its central location, St. Louis, Missouri, has been chosen as the site for the 2011 and 2013 ALS panel meetings, both pilot and operational. The 2011 ALS meetings will involve two grade levels, grades 8 and 12, and the 2013 meeting will involve one grade level, grade 4. Since the 2011 meetings require the more complex design, this Design Document details the studies for grades 8 and 12. The grade 4 studies will follow the same design. There will be 20 panelists per grade for the pilot study and 30 per grade for the operational ALS meeting. In addition, each grade-level panel will have a process facilitator and a content facilitator.

The 2011 ALS implementations will include both whole-group sessions and grade-group sessions. Following registration, the ALS meeting will begin with an orientation session for all panelists. Throughout the ALS meeting, all training and instructions will be given to the whole group. Instructions for specific tasks will be repeated in the grade group where panelists can ask detailed questions relevant to their tasks. All training materials will be loaded in BoWTIE and available to the panelists during their grade-group activities. In addition, the ALS descriptions will always be available in the grade-level rooms. A preliminary agenda is shown in Figure 3.

Figure 3. Agenda for 2011 NAEP Writing ALS Meeting

Agenda	
2011 NAEP Writing for Grades 8 and 12	
Achievement Levels-Setting (ALS) Meeting	
February 7–10, 2012	
St. Louis, Missouri	
Day 1 – February 7, 2012	
8:00 AM	Registration
	<i>Whole Group Session</i>
8:30	Welcome and Introductions
	General Orientation
	NAEP
	NAEP ALS Process
	<i>Grade Group Session</i>
9:30	Panelists' Introductions and Taking a NAEP Exam
10:30	Break
10:45	Scoring and Discussion
	<i>Whole Group Session</i>
11:15	Orientation to the Method
12:15 PM	Lunch
	<i>Whole Group Session</i>
1:15	NAEP Writing Framework Presentation
	NAEP Achievement Levels Presentation
2:45	Break
	<i>Grade Group Session</i>
3:00	NAEP Writing Achievement Level Descriptions (ALDs) Review and Discussion
	Evaluation #1
5:30	Adjourn

Day 2 – February 8, 2012

Whole Group Session

8:00 AM Recap of Previous Day’s Activities

Grade Group Session

Achievement Level Training by Reviewing Prompts and Scoring Rubrics

8:45 KSA Review

10:00 Break

10:15 Discussion for Common Understanding of the ALDs

11:30 Lunch

12:00 PM Training on the Rating Method

Grade Group Session

1:30 Round 1 Ratings (Range-finding)

Evaluation #2

5:30 Adjourn

Day 3 – February 9, 2012

Whole Group Session

8:00 AM Recap of Previous Day's Activities
Round 1 Feedback Explanation

Grade Group Session

9:00 Feedback from Round 1
Cut Scores
Rater Location Feedback
P-Value Feedback (Overall Difficulty of Each Prompt)
Reckase Chart (Conditional Difficulty of Each Prompt)
Review ALDs

10:30 Break

10:45 Review Sample Responses with High Rate of Disagreement Among Panelists' Ratings
Evaluation #3

12:30 PM Lunch

1:30 Round 2 Ratings (Range-finding)

Whole Group Session

4:00 Feedback from Round 2
Cut Scores
Rater Location Feedback
Consequences Data

Grade Group Session

4:30 Feedback from Round 2 and Discussion
Cut Scores
Rater Location Feedback
Consequences Data
Evaluation #4

5:30 Adjourn

Day 4 – February 10, 2012

7:30 AM Breakfast

Grade Group Session

8:00 Round 3 Ratings (Pinpointing)

11:00 Break

Whole Group Session

11:15 Feedback from Round 3

Cut Scores

Rater Location Feedback

Consequences Data

Grade Group Session

11:45 Feedback from Round 3 and Discussion

Cut Scores

Rater Location Feedback

Consequences Data

Consequences Data Questionnaire

12:15 PM Selection of Exemplar Items

Whole Group Session

12:45 Evaluation #5 and Debriefing

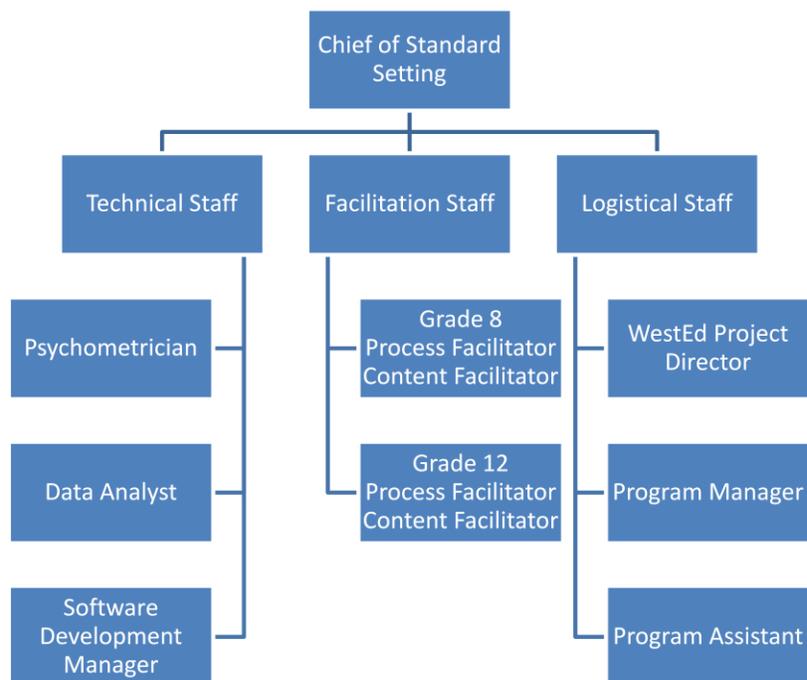
1:15 Adjourn

Facilitation

All whole-group sessions will be facilitated by the chief of standard setting. All grade-group activities will be facilitated by the grade-level facilitator. Each grade-level group will have a content facilitator who has expertise in writing and the NAEP writing framework. Cutpoints and other grade-level information provided after a round of rating will be presented to the whole group. All other feedback will be presented in the grade group following whole-group orientations on the feedback to be provided.

An organizational chart that describes the management of the ALS meeting is in Figure 4. Please note a manager from the Measured Progress software engineering staff will be available onsite to assist with the use of the computer-based tool.

Figure 4. ALS Facilitation Staff



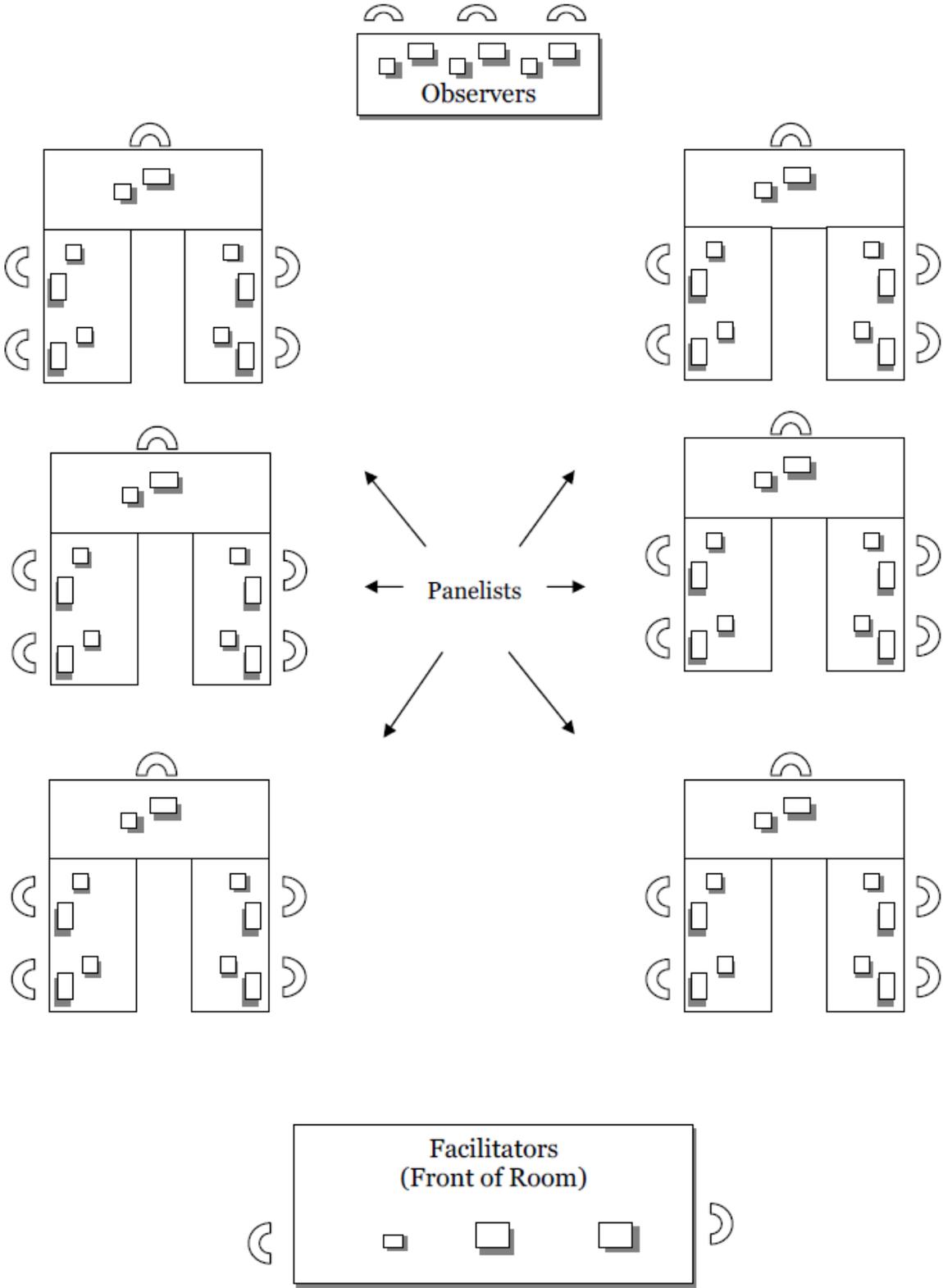
To maximize the extent to which the process is implemented identically across grade levels, the facilitators will be trained by the chief of standard setting. A one- to two-day facilitator training session will be held in Dover, New Hampshire, approximately two weeks before the pilot study meeting. Additionally, a facilitator script will be prepared for use in the meetings. The facilitator script is a step-by-step set of instructions to be used during the entire standard-setting process.

The day before the ALS meeting, the facilitation staff will meet with the chief of standard setting, the COR, and the lead psychometrician to undergo a refresher training. Additionally, at the beginning of each day and prior to the panel meetings, the chief of standard setting will review with the facilitation staff the processes that will be implemented for that day.

Room Setup

Because the 2011 meetings involve two grades, two sets of facilitators, and two groups of panelists, one large meeting room will be used for whole-group activities among the 60 panelists (40 for the pilot study), and separate smaller meeting rooms will be used for grade-level activities. The room layout is designed to provide enough space to accommodate a NAEP laptop and a netbook for each panelist with easy access to wiring. Figure 5 illustrates a typical room setup for the grade-level meeting rooms.

Figure 5. Room Configuration for ALS Process for NAEP Writing 2011/2013



Panelist Training

Panelist training is designed to prepare panelists to properly perform their tasks. Training is also designed to assure that panelists understand the BoW procedures and writing NAEP and feel comfortable about the training and instructions. Sufficient time for training is designed to enhance procedural validity and the ability of panelists to make informed judgments that result in achievement levels that are reasonable, valid, and informative to the public. The panelist training described here is consistent with the training provided to the panelists for the 1998 NAEP ALS meetings, which exemplifies a thorough training program for standard-setting panelists (Raymond & Reid, 2001). Modifications will be made to adjust training to address needs specific to the rating method. BoWTIE training is included in each aspect of the implementation process.

General Orientation

Onsite training begins with a general orientation to the NAEP program and the role of the National Assessment Governing Board. An overview of the NAEP program is presented by the Governing Board COR. This is followed by the general introduction to the NAEP ALS process, emphasizing the steps related to the overall process that have already taken place and the steps that will follow after the conclusion of the ALS meeting. The intent is to provide the panelists as much context as possible, so that they are well informed when they get to the rating task. This type of information is provided to all the panelists at the same time, ensuring grade-level panels are provided the same training to the greatest extent possible.

Taking a NAEP Exam

Given that the goal of the ALS process is to determine what students should know and be able to do, it is logical for the panelists to become familiar with how students experience the test. Early in the process, each panelist will take a form of the NAEP at the grade level for which he or she will be setting the cutpoints. This step is performed in the grade-level group. This will be the panelists' first exposure to the NAEP assessment. Panelists will be given a brief orientation to the test booklet and the test-taking situation. Since the 2011 writing NAEP will be administered by computer, panelists also will take the assessment on the computer. Through Westat, NCES will make computers available for the panelist to take the test, and these will be the computers used to administer the writing NAEP in 2011. (This is a different computer than the one they will use to access BoWTIE.) On-site technical support for the test-taking computers will be provided by the contractor for computer-based administration of 2011 NAEP writing. The test-taking computer will also be used by the panelists to view the other prompts in the assessment. Each panelist will score his or her own completed test using the same scoring rubrics employed by the operational scorers. They will be instructed to review their responses using scoring guides, and they will be told that their tests will not be scored or used in any other way.

Orientation to the Method

Panelists will again come together as a group to train in the standard-setting method. Panelists will be given an overview of the BoW method as well as an overview of the steps in the process, such as the rounds of ratings and the feedback provided after each round. The goal is to provide the panelists the rationale for all steps to help

them understand their task without distracting panelists by providing too much information.

Presentation of NAEP Writing Framework and ALDs

Just as it is important that the panelists understand the assessment, it is important that they understand its framework. The framework is the ultimate source of information about the assessment. A whole-group presentation on the framework will be made by the content facilitators who were selected from among the persons who developed the framework and who work with the NAEP writing program for item development and review. The goal of the presentation is to inform the panelists about the framework and achievement levels in a manner that will contribute to the confidence of the panelists and the integrity of the process.

The training on the ALDs will continue in each grade-level group where the content facilitator will lead a discussion and allow panelists to ask questions specific to their grade level. After the discussion, they will be ready for exercises aimed at familiarizing panelists with the writing prompts in the assessment and gaining a common understanding of the ALDs.

Achievement Level Training by Reviewing Prompts and Scoring Rubrics

After panelists have been introduced to the writing assessment, frameworks, and have had discussions of ALDs, they will review all prompts and scoring rubrics for their grade level. Again, given that the assessment was administered to students on a computer, panelists will only be able to access the prompts through the NAEP administration computer.

In order for the grade-level panelists to provide ratings that will yield a set of reliable and valid cutpoints, it is imperative that they gain a common understanding of the ALDs. In the absence of a collectively shared understanding of what students should know and be able to do, the cutpoints resulting from the process will, more than likely, have no valid interpretation. Therefore, the next set of activities is geared toward gaining common understanding of the ALDs.

Panelists will work to develop a list of the knowledge, skills, and abilities (KSAs) needed for each score level for each prompt based on its scoring rubric. The KSAs for each score level will then be compared to each achievement level. A discussion of these comparisons will be instrumental in gaining a common understanding of the ALDs. Although there will be no modifications to the ALDs, functionalities of BoWTIE will allow panelists to highlight text and mark annotations on the ALDs so they can remember the meaning of each description as they go through the rating process.

Training on the Rating Method

Within each grade group, a sample of student work selected from the prompts that are common to both rating groups will be used to train panelists in the rating method. The sample presented will be rank-ordered from lowest to highest level of performance, but the panelists will not know the specific score for each booklet or prompt. For each of the sample responses, the panelists will determine, as a group, which achievement level corresponds to the KSAs exhibited in the response. Based on the KSAs exhibited, panelists will decide as a group the achievement level to which it will be classified. Panelists will be instructed on how to use BoWTIE for accessing the responses and on providing their ratings based on the achievement level classifications.

Rounds of Rating and Feedback Information

Panelists will be introduced to the ratings task as a whole group but will be given more specific instructions in their smaller grade groups. Data will be collected for each round of ratings and will be analyzed so that results and feedback information can be given to the panelists to inform their ratings for the subsequent rounds. Feedback information will be given to the panelists to guide their judgments and provide indications of how well they are performing their task. Further, feedback “provides evidence for the quality of the conduct of the process as well as a direct indication that the standard setters considered relevant information when participating in the process” (Reckase, 2001a).

To discourage the panelists from comparing their results with the results of the other grade group, a pseudo-NAEP scale will be developed for each of the groups. The two scales will be different linear transformations of the NAEP scale.

Round 1: Rangefinding

During this round, panelists examine sets of work samples that are distributed across the full score range. The work samples will be presented on the computer in order of performance from lowest to highest based on their EAP scores. Annotation functionality will be available so that panelists can make notes regarding each student work sample. The panelists will be asked to rate the student work sample and input their ratings on the computer. The rating procedure involves categorizing the student work sample into an achievement level of Basic, Proficient, or Advanced. If the student work sample does not meet the criteria for being placed into any of these categories, the

work sample is categorized as Below Basic. Panelists will have an opportunity to review their ratings before submitting them for analysis.

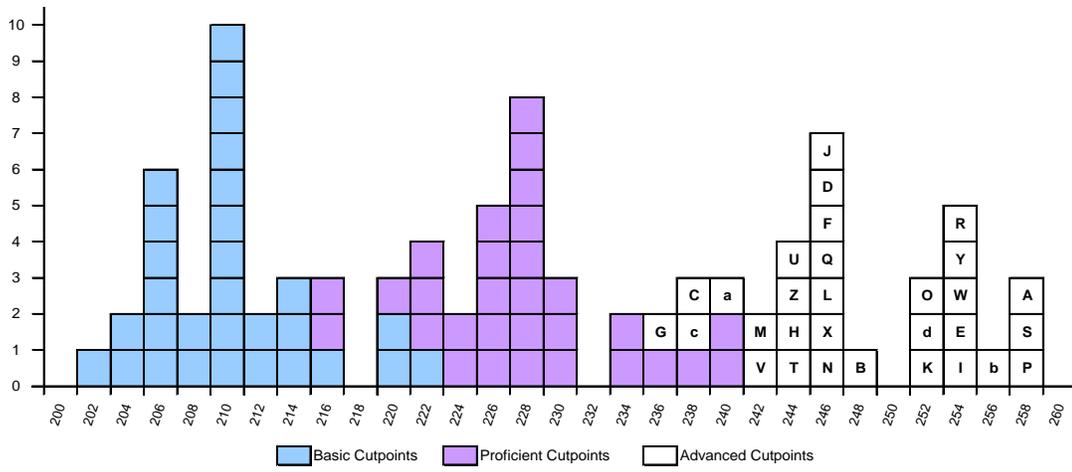
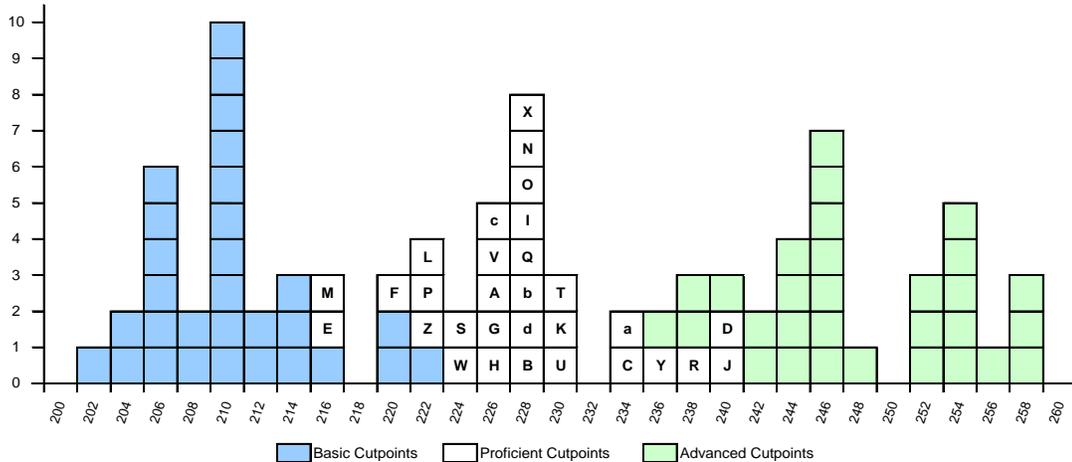
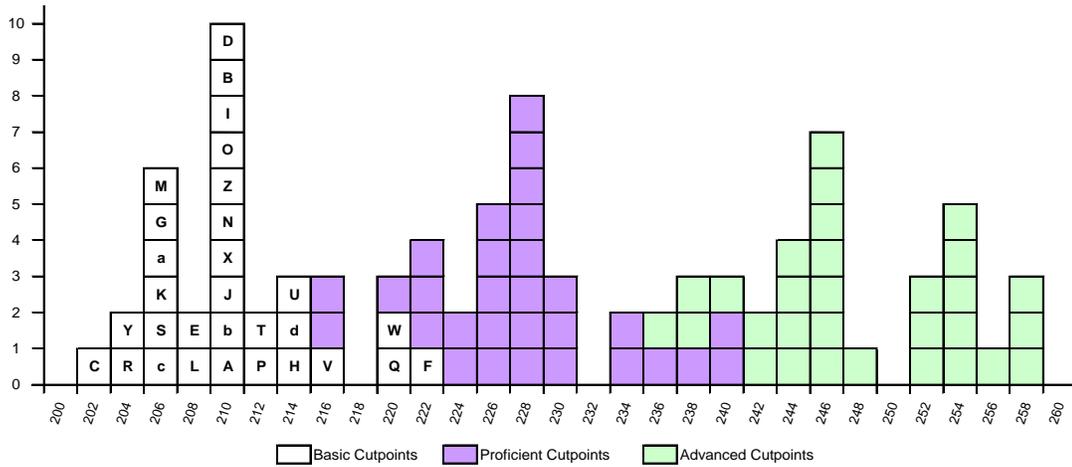
Feedback from Round 1

Grade-level cutpoints, standard errors, rater location charts, and the Reckase Chart will be presented to inform panelists' second round of ratings. The grade-level cutpoints will be presented to panelists by the chief of standard setting to inform them about the work of each grade-level group. In cases where there are large differences in specific cutpoints across grades, panelists will be cautioned not to over-interpret such results given that the writing NAEP scales are independent across grade levels.

Rater Location Charts

The rater location charts graphically present the distribution of cutpoints set by panelists, where each panelist's cutpoint is indicated by a letter code known only to him or her. One rater location chart will be produced for each achievement level. A chart for a particular level identifies panelists' cutpoints for that level, while the cutpoints for other levels are not identified by code. This chart is used to inform the panelists about where they set their cutpoints and whether the cutpoints were set higher, lower, or similar to those set by other panelists. An example is provided in Figure 6 below.

Figure 6. Rater Location Feedback



P-Value Feedback

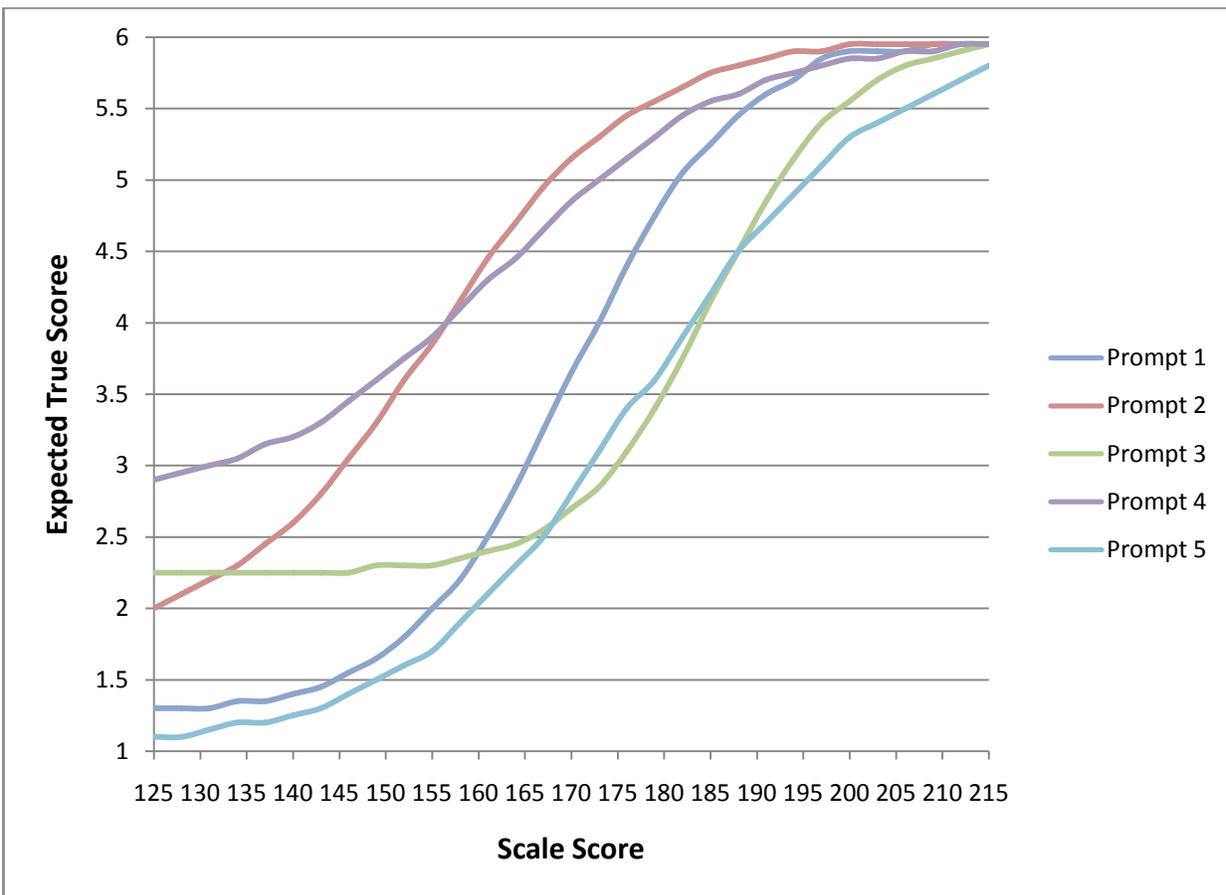
The p-value data provide a measure of actual student performance by reporting the mean score for each writing prompt for all of the students who took the NAEP test. Important as a reality check, the p-values indicate the level of difficulty for each writing prompt. Given two writing prompts, the prompt with a higher p-value is easier than the writing prompt with a lower p-value since a greater number of students achieved a higher score on the writing prompt with a higher p-value. The p-value feedback also includes the percentage of students who performed at each score level.

Reckase Charts

The Reckase Chart is a feedback mechanism to show how panelists' ratings relate to the performance of students on the writing prompts. The Reckase Chart shows the relative difficulty of the writing prompts and the rate at which the performance on the writing prompts changes as the performance of examinees increase on the pseudo-NAEP scale. Figure 7 shows an abbreviated graphical example of a Reckase Chart with five prompts. The horizontal axis is the pseudo-NAEP scale. This scale is purposely different from the NAEP score scale used for reporting so that panelists will not know the official NAEP results before the results are approved for release by the Governing Board. The vertical axis is the expected score of examinees who have a particular scale score. For example, for the examinees estimated to have a scaled score of 152, the expected average score would be 3.75 on Prompt 4. Prompt 5 is substantially harder for these students because their expected score on this prompt is 1.6. In contrast to the p-value data that provides information on the overall difficulty of each prompt, the

Reckase Chart provides the conditional difficulty of each writing prompt. Both types of prompt difficulty feedback are provided to the panelists for a reality check.

Figure 7. Example of a Reckase Chart



Round 2: Rangefinding

During the second round of ratings, panelists will be presented with the same set of student work, along with the classifications they provided in round 1. Their task is to provide an achievement level classification for each student work sample, similar to what they did in round 1, in light of new information as well as feedback from the first round of ratings. This may be viewed as an adjustment to the rating provided in round 1.

Prior to performing the rating task, the panel will review the ALDs and engage in a discussion of the work samples for which their ratings were most disparate. The criteria for selecting the work samples to discuss will be based on both the rate of disagreement and the number of work samples that can be reviewed within the time allocated. The first part of the discussion will be for the grade group led by the process facilitator with the content facilitator providing expertise as needed. Only student work samples from the common forms will be eligible for grade group discussion. It is expected that the discussions will yield descriptions of what differentiates levels of performance between adjacent achievement levels. After discussing student work samples from the common forms, discussions will continue within each table group of student work samples from forms unique to the group. Discussions will be about student work samples for which panelists have a high rate of disagreement or about specific student work samples that particular panelists would like to discuss with other panelists at the table.

Feedback from Round 2

Ratings from round 2 will yield new cutpoints and standard deviations and will be used to produce new rater location charts. This feedback will again be presented to the whole group and will be discussed in the grade group prior to the second round of ratings. Additional feedback from round 2 results will be the consequences data.

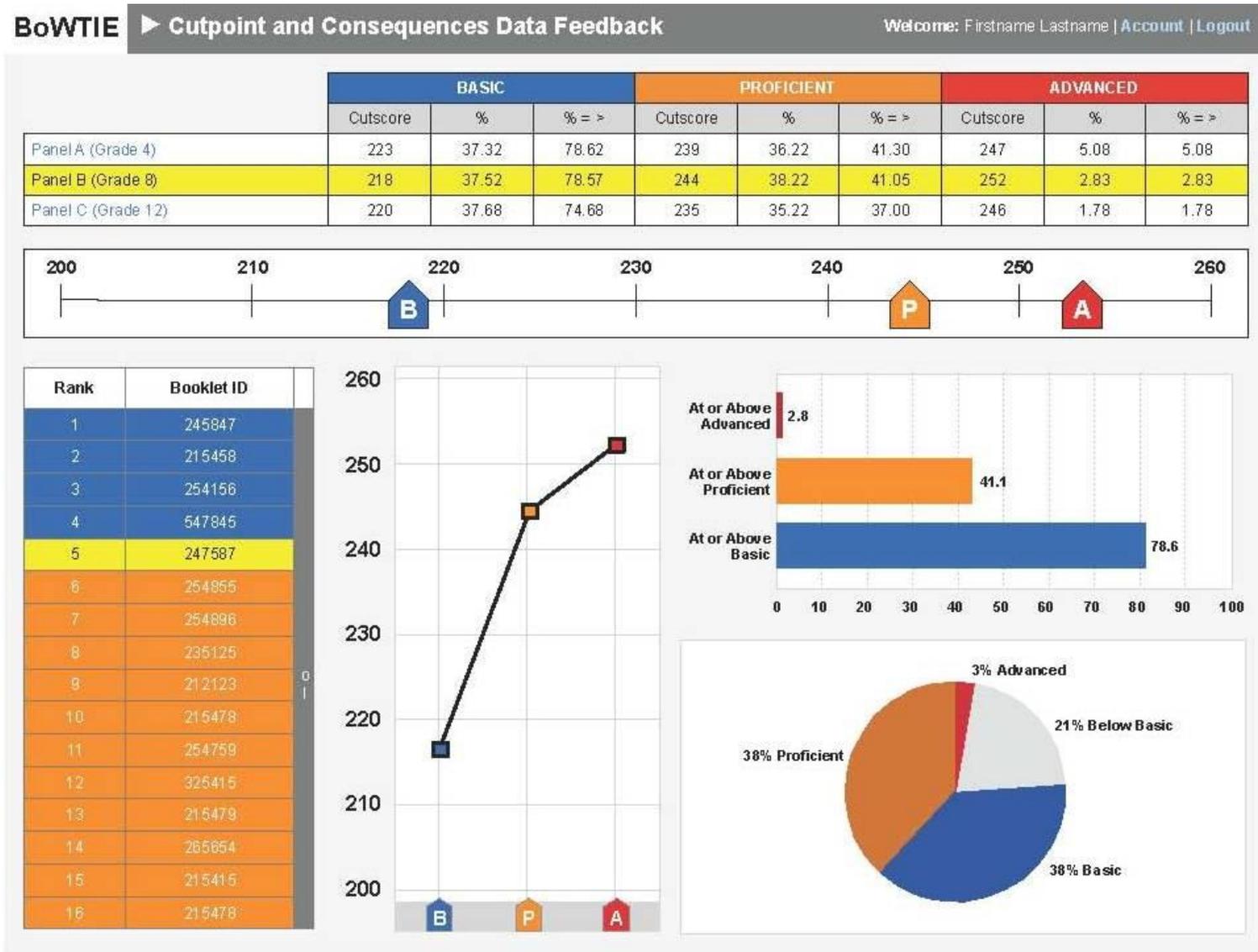
The consequences data feedback informs the panelists of the proportion of students that would fall at or above the cutpoint of each achievement level, based on their grade-level cutpoints. If the proportion of examinees does not match a panelist's expectations, based on the ALDs and their own experience with students, they should

reexamine their ratings relative to the ALDs and determine if adjustments are needed. Often only slight adjustments are needed for panelists to become comfortable with their ratings.

Consistent with Cizek and Bunch's (2007) sentiment with moving cut scores, such action would only be justified on the basis of student performance at that cut score. BoWTIE includes an interactive feedback mechanism for the consequences data feedback. The wireframe for this mechanism is in Figure 8.

The consequences data feedback software is an interactive tool that provides the panelists easy access to student work samples when they move the cutpoints. It also will instantaneously provide the new consequences data resulting from changing cutpoints. The software will calculate the percentage of students at each scale score as well as highlight the student work sample to which the scale score is attached.

Figure 8. Consequences Data Feedback



The Cutpoint and Consequences Data Feedback screen can be accessed by each individual panelist using a netbook computer, and it can be used by the facilitator to lead a group discussion. At the top of the screen is a table of cut scores and consequences data with three rows—one for each grade level. Immediately below this table is an interactive slider. To use the slider to move the cutpoint for one of the grade levels, the row for the desired grade level must be highlighted. Once a row is highlighted for grade 4, 8, or 12, moving or sliding the cutpoint will cause the following real-time changes:

- The cutpoints themselves will change, along with the percentage of students scoring at or above that level.
- Changing the cutpoint will change the percentages of students scoring in the adjacent levels. That is, the percentages of students in the Basic and Proficient levels both change when one moves the Proficient cutpoint. Changes in the percentage will be seen on the table of numerical results for the highlighted grade and on the bar graph. Changes to the percent of students in each achievement level will be seen on the pie chart.
- Changing the cutpoint will result in updated information in the table of numerical results and on the line graph.
- Changing the cutpoint will also change the highlighted student work sample displayed in the two-column table vertical slider on the left-hand side of the screen. On the vertical slider, the scale and work sample move relative to the highlight when the cut score is moved. Easy access to the work samples at or

around the cutpoints will help the panelists maintain the necessary connection between the cutpoints and the ALDs.

Round 3: Pinpointing

Based on cutpoints computed from the latest rangefinding round, sample booklets will be selected with scores around the cutpoint. These samples will be the booklets that panelists will classify into achievement levels in the pinpointing round. The task will be similar to the task for round 1 except panelists will have a new set of student work samples to classify into achievement levels. Panelists will be provided sample student work from each form in their rating pool.

For each cutpoint, panelists will be presented 15 student work samples with EAP scores around that cutpoint. The lowest score in the sample will be lower than the lowest individual cutpoint set by a panelist in the grade group. Similarly, the highest score in the sample will be higher than the highest individual cutpoint set by a panelist in the grade group. The scores of the 15 student work samples will be uniformly distributed within the specified range. Unlike the rangefinding booklets, pinpointing booklets will not be presented in rank order according to score. The panelists' task will be to classify each booklet below or above the cutpoint based on their understanding of the ALDs and the level of performance exhibited by the student. The classification task will be performed separately for each cutpoint.

Feedback from Round 3

Feedback from round 3 includes grade-level cutpoints and a standard error computed from panelists' ratings. Round 3 feedback also includes rater-location data

based on individual cutpoints. The consequences data will again be provided as feedback from round 3.

Consequences Data Questionnaire

After the presentation and discussion of consequence data from round 3, panelists will fill out a questionnaire to indicate whether they want to make additional changes to any of the cut scores after learning the consequences of those cut scores. Panelists will be able to recommend a change for any or all three cut scores. The questionnaire will help Measured Progress provide more information to the Governing Board to inform their policy decision on the cut scores.

Selection of Exemplar Items

Exemplar items are one of the products of the ALS process and one of the three parts of the NAEP achievement levels. These items exemplify what students know and can do at each of the achievement levels. Exemplar items are selected from a set of items that are to be released to the public. For the 2011 writing NAEP, the plan is to release one prompt for each writing purpose.

Each writing prompt is scored using a six-level rubric, where each score level requires a higher level of performance than any score that is lower. In the context of NAEP writing, an item is a combination of a writing prompt and a partial- or full-credit score. Thus, for each grade level there are effectively 15 items from which to select.

There are two stages for selecting which items exemplify performance at each achievement level. First, statistical criteria will be applied to determine which items are eligible for selection as exemplars for each level. In the second stage, panelists, based on their understanding of the ALDs, will provide a recommendation for which items

exemplify what students can do at each achievement level. For each achievement level, panelists will be presented with the list of potential exemplar items. For each of these items, they will also be provided with sample student responses to provide a point of reference to help them make their selection. These sample responses will be clear examples of student responses that received a score corresponding to the rubric level for which an exemplar item is being considered.

They will be instructed to record their initial ratings. Then for each item they will be asked to indicate if they recommend the item as an exemplar for that level. Each item for which there is a disagreement will be discussed as a group. Consensus is a goal of the discussion, but not a requirement. Once the items have been discussed, panelists will provide their final recommendations by filling out a survey. Both sets of recommendations will be summarized and included in the reports to the Governing Board.

Evaluation

At the end of the first day and after each round, panelists will be provided with an evaluation form designed to assess their understanding of instructions, tasks, and materials. Five questionnaires are planned for administration over the course of the panel meetings. The schedule of the five evaluations is described in Table 14 below.

Table 14. Schedule of Process Evaluations

Evaluation	Schedule
#1	End of Day 1
#2	End of Day 2 after Round 1
#3	Before Lunch on Day 3 after Round 1 Feedback
#4	End of Day 3 after Round 2 Feedback
#5	End of Day 4 after Round 3 Feedback and before Debriefing

Most panelist responses to the evaluations will be collected on Likert scales, but several responses will be narratives that address specific aspects of the process. These evaluations will be reviewed at the end of each day and any sources of confusion, dissatisfaction, or other concerns will be identified for clarification with individual panelists or the panel as a whole. Summary results of each evaluation questionnaire are made available to the facilitation staff as well as the COR shortly after all the data are collected for each questionnaire.

Security Procedures for Meetings

For the ALS process, netbook computers and servers will be mailed to meeting locations with no data in the storage drives. Measured Progress personnel will carry the data to be used for standard setting to the site where a local hosting option will be used. After the meeting, a Measured Progress software engineer will purge each netbook computer and server prior to mailing them back to Measured Progress. Measured Progress personnel will hand-carry the external hard drives to Measured Progress.

Additionally, the test administration contractor will provide secure transport for the NAEP laptops.

Each panelist will be asked to sign the compliance form as required by NCES. Security measures for meetings will also require that an onsite office or workroom be set up for the pilot study and ALS meetings, and that room must be secure at all times. This means imposing a 24-hour hold on the given room. Hotel personnel will not have access without the presence of an authorized Measured Progress employee. In the event that secure materials need to be left in the meeting rooms, ALS staff will have the key to lock and unlock the rooms to maintain security. At the end of each day, all materials will be collected, accounted for, and stored in the workroom.

Another aspect of NAEP security is an assurance that results of the ALS meetings are not prematurely released. As is traditional in the NAEP ALS process, the NAEP scale will not be used to provide feedback to panelists. Instead, a NAEP-like scale obtained through linear transformation from the NAEP scale will be used for this purpose.

Public Comment Forums

A crucial aspect of this work will be to obtain input from a wide variety of sources and provide an opportunity for ongoing feedback to guide our work. Specifically, WestEd will design and implement three public comment initiatives: one focused on receiving input from a broad spectrum of groups and individuals regarding the proposed design for the writing ALS and procedures; one focused on receiving feedback on the writing ALS outcomes for grades 8 and 12; and one focused on receiving feedback on the writing ALS outcomes for grade 4. The three public comment

initiatives fit into the overall process for ensuring that all aspects of the writing ALS process receive technical guidance, Governing Board approval, and input from a wide range of constituents.

Public Comment Initiative #1

Soliciting public comment on the ALS design, the first public comment initiative will be integral to establishing the ultimate procedural validity and field consensus for the final ALS design and procedures. This public comment initiative will coincide with the development of the ALS design: solicitation of public comment will commence on February 9, 2011, and will conclude on February 24, 2011—prior to Measured Progress’s submission of the final draft of the Design Document to the Governing Board. This schedule will allow for the most complete version of the Design Document to be reviewed by the public, while allowing time for Measured Progress to make necessary modifications to the Design Document prior to final draft submission. The intent of this initiative is to obtain comment from a broad spectrum of educators, policymakers, and the general public, including to the greatest extent possible the following:

- Key organizations that collaborated in the review of the 2011 Writing Framework: the Association for Supervision and Curriculum Directors (ASCD), the Conference on College Composition and Communication (CCCC), the Conference on English Leadership (CEL), the Council of Chief State School Officers (CCSSO), the Council of Great City Schools (CGCS), the International Reading Association (IRA), the National Association of Elementary School Principals (NAESP), the National Council of Teachers of

English (NCTE), the National Writing Project (NWP), and Teachers of English to Speakers of Other Languages (TESOL)

- Common Core Standards consortia and key stakeholders: the SMARTER Balanced Assessment Consortium, the Partnership for the Assessment of Readiness for College and Careers, and the Validation Committee for the Common Core State Standards
- Other relevant organizations, such as the National Education Writers Association (National EWA), the National Association of Secondary School Principals (NASSP), the International Society for Technology in Education (ISTE), and the National Association of Private Catholic and Independent Schools (NAPCIS)

Notification of the public comment solicitation will be posted on the Federal Register, the Governing Board website, the Measured Progress website, and the WestEd website. It will also be communicated via social networking media (i.e., the WestEd Facebook page and Twitter). The notification will provide background and context for the initiative and encourage individuals to access a webpage, to be hosted on WestEd's server, that will reiterate the objective of the initiative and provide access to the Design Document and guiding questions. Individuals will be directed to submit comments and/or recommendations to Jennae Bulat at WestEd.

Guiding questions for comment follow. Panelists may respond to any or all of these questions, and they may comment on other issues not addressed in these questions.

1. The objective of this study is to set achievement levels for the 2011 and 2013 NAEP writing assessments. Does the study design as presented in the Design Document seem reasonable for accomplishing this overall study objective?
2. What improvements can be made to the design to more fully accomplish the objectives of this study?
3. The proposed design calls for the computerization of many aspects of the study. Are there aspects of this computerization that will be particularly effective or ineffective in meeting the objective of this study?
4. Is the field trial as described a reasonable method for testing the computerization of the methodology?
5. Is the special study as described a reasonable method for comparing performance on the 2007 writing NAEP assessment with performance on the new writing NAEP assessment?

Results will be collected and recommendations categorized, to the extent possible, so that they can be evaluated more easily. The recommendations will then be presented to the TACSS for consideration and recommendations. Any recommended changes in the design will be presented to the Governing Board for consideration and approval.

Public Comment Initiatives #2 and #3

The second and third public comment initiatives will be tied to the proposed achievement levels for grades 8 and 12 and for grade 4. Public comment on the achievement levels will serve as a critical step in gaining valuable feedback from concerned constituents regarding the face validity of the achievement levels, their

policy implications and expected and unexpected consequences, and the potential variance across the nation in terms of writing achievement level expectations for grades 4, 8, and 12. These public comment initiatives will precede the reporting of panel-recommended achievement levels to the Governing Board. Solicitation of public comment for grades 8 and 12 will commence on or about March 12, 2012, and solicitation of public comment for grade 4 will commence on or about March 11, 2014, with each session lasting approximately two weeks.

Because of the need to maintain the confidentiality and security of NAEP data prior to the release of the Nation’s Report Card by the Commissioner of Education Statistics, it will be necessary to limit public comment for these two initiatives to individuals from whom signed confidentiality agreements can be obtained. For this reason, we propose hosting a series of focus groups. Working with the Governing Board, WestEd and Measured Progress will recommend four towns/cities—one from each NAEP region—that represent a range of size, urbanicity, and SES status. In each location, WestEd will conduct 2–3 focus groups of 5–8 members each. For each focus group, WestEd will recruit members of the organizations and groups listed under Initiative #1 as well as local educators, parents, policymakers, and members of the general public, all of whom will be familiar with the writing skills and abilities of students at the relevant grade level.

These groups will be convened in separate two-hour sessions. Sessions will be held in WestEd/Measured Progress/EPIC facilities if possible; if such facilities are not available, local hotel meeting rooms will be secured. Prior to each focus group, participants will be sent NAEP policy definitions and achievement level descriptions to

review. At each focus group, participants will be required to sign confidentiality agreements, after which the group will be shown the achievement level cut scores recommended by ALS panelists, consequences data (i.e., percentages at and above the achievement levels recommended by ALS panelists), exemplar items recommended by ALS panelists, and sample student responses to these exemplar items.

The group will then be encouraged to provide feedback on the following guiding questions:

- How reasonable are the percentages at and above each of the achievement levels, given your knowledge of writing and the descriptions of the NAEP writing achievement levels?
- How useful are the illustrative responses in communicating what students should be able to do in writing at each achievement level for each grade?

In addition, participants will be allowed to express additional comments, recommendations, and/or concerns as desired either during the focus groups or after the groups via email directed to WestEd.

Focus group responses will be summarized to identify trends in responses, with specific comments evaluated on their merit. Relevant results will be communicated to the Governing Board and used to modify the ALS design and procedures as deemed necessary.

Reporting of Results

Two sets (grades 8 and 12, and grade 4) of comprehensive final reports composed of a Process Report and a Technical Report will be produced at the conclusion of the ALS meetings. The Process Report will include, but not be limited to:

- A Description of the ALS Process—The procedural aspects of the entire ALS process, including the field trial and pilot study, will be documented by Measured Progress in a way consistent with providing evidence of procedural validity which is discussed in the Validity Evidence section of this document. The Process Report will document the elements of the ALS process and panelists’ reactions to each. The extent to which this information provides evidence of procedural validity will be evaluated and reported along with the extent to which the Design Document was implemented. We will include data on panelists’ ratings, changes in ratings, internal consistency of ratings, panelists’ reactions to feedback data, and panelists’ confidence in the process and satisfaction with the process and process outcomes. The report will include descriptions of participants, training, meetings, tasks, materials, results, and final recommendations.
- Recommended Achievement Levels—Three recommended levels—Basic, Proficient, and Advanced—will be presented. All relevant data obtained during the ALS process will be aggregated for presentation to the Governing Board, including appropriate summary statistics to include median and average cutpoints, standard deviations of those cutpoints, percent of examinees scoring at or above each achievement level, and a comparison of cutpoints and impact data across subgroups; estimates of reliability; and confidence intervals. Additionally, achievement-level data for groups and subgroups used in NAEP reports will be included in briefings to the Governing Board.

- **Achievement Levels Descriptions and Exemplars—Measured Progress**
acknowledges that clear and coherent descriptions of what students should know and be able to do at each achievement level for writing for grades 4, 8, and 12 will be developed by the Governing Board and provided for use throughout the ALS process. The process for selecting exemplar items developed for the 2011 and 2013 NAEP in writing will be described. The exemplar items will be part of the achievement levels that will be recommended to the Governing Board. The relationship and alignment to the achievement levels definitions and the scale score range will be made explicit.
- **Public Comment—Procedures and results related to obtaining public comment on the design of the process will be summarized and presented to the Governing Board in the final report.**
- **Recommendations for Future ALS Activities—Based on the levels-setting activities conducted under this procurement and previous procurements, Measured Progress will provide a thoughtful and reflective discussion of recommendations for future ALS efforts.**

The Technical Report will include, but not be limited to, the complete documentation of the ALS aspects delineated below. It will be not only clear, concise, and complete, but also comprehensible for all interested persons, including persons who are not trained in educational measurement.

- **Technical Advice—All technical advice and decisions reached during this project, along with rationales and considerations for the decisions, will be**

documented and summarized. In addition, a complete set of minutes for the TACSS meetings will be appended to the report.

- **Data Analysis Procedures—Measured Progress** acknowledges that all analysis techniques, decisions, formulas, and procedures must be documented clearly and completely. Additionally, we understand that original data forms, software created under the contract, and all raw data will become the property of the Governing Board and will not be released by the contractor to any other parties without written consent from the Governing Board.
- **Pilot Activities—**Technical aspects of the pilot activities conducted as part of the level-setting process will be described and documented in this report, including data collection activities, data analysis, and results.
- **Materials Analysis and Description—**All materials used throughout the process, including briefing booklets, agendas, prompt pairings and prompt assignments to panelists, and other relevant information, will be described and documented.

Validity Evidence

In an endeavor that relies primarily on value judgment, validity of results relies primarily on procedural and internal evidence. Procedural validity stems from evidence indicating that procedures are reasonable, were carried out as intended, and were understood by panelists. Internal validity stems from evidence centered on comparisons of results using the same methods in different occasions, variability of each panelist’s cut scores across rounds, variability of cut scores among panelists, and variability of cut scores across rating groups. A collection of evidence documenting the

procedural and internal validity of the ALS meeting will be included in the final report submitted to the Governing Board. Further, the accuracy and consistency of classification of student performance will be reported as validity evidence for inferences that can be generated from the resulting cut scores.

Procedural Validity

Procedural evidence of validity is the degree to which the entire ALS process is tightly interwoven with strong connections between every component part (Reckase, 2001b). Measured Progress will collect procedural evidence by responding to each of the following questions:

- Was every part of the Design Document reviewed by stakeholder groups?
- Were the processes and procedures executed as planned in the Design Document?
- Was the ALS process fully documented?

The role of documentation of the ALS process in establishing evidence of procedural validity cannot be overemphasized. For the NAEP ALS process, such documentation should, at minimum, include:

- Definition and purpose of the ALS process
- Definition of achievement levels that were used in the process
- A description of the ALS method and rationale for its choice
- Recruitment, selection, and training of panelists
- Feedback from panelists about their understanding of the purpose of the ALS process and their particular task as it relates to the purpose, as well as about their level of satisfaction with the process

- Description of scoring procedures, including the rubrics that were used, that is pertinent to panelists' rating tasks
- Description of procedures applied to compute cut scores for each achievement level
- Description of any study performed to establish technical quality of the assessment in order to compile evidence of validity

Although procedural evidence does not guarantee validity of ALS results, the lack of procedural evidence can negatively affect credibility of results. That is, the lack of validity evidence makes one question the appropriateness or correctness of inferences, decisions, or descriptions made about individuals, groups, or institutions based on ALS results.

Internal Validity

Capitalizing on the current design, internal validity of the cut scores can be measured using two approaches. The first approach compares results when the same methods and materials across different standard-setting meetings are used (pilot versus operational). The second approach compares results when the same procedures with different groups of panelists and different writing prompt sets are used. With both approaches, two sets of cut scores will be obtained and compared. When similar cut scores are obtained, this suggests that the procedures yield valid and reliable cut scores.

After the standard-setting meetings, cut scores will be calculated for various panelist subgroups (e.g., teachers, nonteacher educators, and general public). The extent to which cut scores for the various subgroups are consistent is also an indication that the results are internally valid. It is important to note that for any given set of

subgroups there will be some natural variability in the cut score. To address whether variability is expected, two methods will be employed: 1) a simple bootstrap and 2) an empirical review. The simple bootstrap study will be used to find what the variability in cut scores might be for any random subgroup of panelists. The empirical review will consist of comparing the variability of cut score recommendations by subgroups of panelists provided for the 2007 writing NAEP ALS to the variability of cut score recommendations provided in the current ALS. Comparisons between the variability of subgroups of interest for the 2007 and 2011 writing NAEP ALS and the bootstrap variability will serve as additional lines of evidence to support claims of validity regarding the cut scores.

Finally, because of the extensive training on ALDs the panelists will focus their understanding on what students (for a given cut score) should know and be able to do. Thus, from one round to the next there should be less variability among the panelists in the location of the cut scores. This confirmatory approach will be further evidence of internal validity.

Decision Accuracy and Consistency

The validity of inferences to be made based on the results of standard setting can be empirically evaluated using decision accuracy and consistency indices. Livingston and Lewis (1995) describe one of the several approaches that may be used. With this approach, an error classification is determined for each cut score. Policymakers often investigate whether the final cut scores should be slightly higher or lower relative to the cut score resulting from standard setting.

Within the last several years, techniques to determine decision accuracy and consistency of classifications have been developed for different assessment situations (e.g., Center for Advanced Studies in Measurement and Assessment (CASMA) Research Reports Nos. 7, 9, 13, 22, and 27). Measured Progress will investigate methods for determining accuracy and consistency of classification that are the most appropriate for NAEP writing. Information regarding the accuracy and consistency of classification at different cut scores will be provided to assist the Governing Board in setting final achievement levels.

References

- Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research*, 56, 137–172.
- Center for Advanced Studies in Measurement and Assessment. (2010). Retrieved 7 February, 2011. www.education.uiowa.edu/casma/research_reports.htm
- Cizek, G. J. (Ed.) (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum.
- Cizek, G. J. & Bunch, M.B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.
- Council of the Great City Schools. (2011). Retrieved 4 February, 2011. www.cgcs.org/about/fact_sheet.aspx
- Hambleton, R. K. & Pitoniak, M. J. (2006). Setting performance standards. In R. Brennan (Ed.), *Educational measurement*, 4th ed. (pp. 433–470). Washington, DC: American Council on Education and Westport, CT: Praeger Publishers.
- Kingston, N.M., Kahl, S.R., Sweeney, K.P., & Bay, L. (2001). Setting performance standards using the Body of Work method. In G.J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 219–248). Mahwah, NJ: Erlbaum.
- Livingston, S. A. & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179–197.
- Radwan, N. & Rogers, W.T. (2006). A critical analysis of the Body of Work Method for setting cut-scores. *The Alberta Journal of Educational Research*, 52 (1), 65–79.
- Raymond, M. R. & Reid, J. B. (2001). Who made thee a judge? Selecting and training participants for standard setting. In G. J. Clark (Ed.), *Setting performance standards: Concepts, methods and perspectives* (pp. 119-157). Mahwah, NJ: Lawrence Erlbaum.
- Reckase, M.D. (2001a). Innovative methods for helping standard-setting participants to perform their task: The role of feedback regarding consistency, accuracy, and impact. In G.J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 159–174). Mahwah, NJ: Lawrence Erlbaum.

- Reckase, M. D. (2001b). The controversy over the National Assessment Governing Board Standards. In D. Ravitch (Ed.), *Brookings papers on educational policy 2001* (pp. 231–265). Washington, DC: Brookings Institution Press.
- United States. Census Bureau. (2011, January). *Poverty*. Retrieved January 14, 2011 from the U.S. Census Bureau website:
www.census.gov/hhes/www/poverty/about/overview/index.html
- United States. Department of Education. National Center for Education Statistics. Institute of Education Sciences. (2010a, August). *Common Core of Data*. Retrieved February 4, 2011 from the National Center for Education Statistics website: <http://nces.ed.gov/ccd/ccdata.asp>
- United States. Department of Education. National Center for Education Statistics. Institute of Education Sciences. (2010b, August). *Local Education Agency (School District) Universe Survey Data*. Retrieved February 4, 2011 from the National Center for Education Statistics website:
<http://nces.ed.gov/ccd/pubagency.asp>
- United States. Department of Education. National Center for Education Statistics. Institute of Education Sciences. (2010c, August). *Private School Universe Survey Data*. Retrieved February 4, 2011 from the National Center for Education Statistics website: <http://nces.ed.gov/surveys/pss/pssdata.asp>
- United States. Department of Education. National Center for Education Statistics. Institute of Education Sciences. (2010d, August) *Public Elementary/Secondary School Universe Survey Data*. Retrieved February 4, 2011 from the National Center for Education Statistics website: <http://nces.ed.gov/ccd/pubschuniv.asp>
- United States. Department of Education. National Center for Education Statistics. Institute of Education Sciences. (2011, November 9). *Interpreting Reading Results: NAEP Reporting Groups*. Retrieved 22 December, 2010 from the National Assessment of Educational Progress website:
<http://nces.ed.gov/nationsreportcard/reading/interpret-results.asp#repgroups>

Appendices

Appendix A

Members of the Technical Advisory Committee for Standard Setting (TACSS) for the 2011 and 2013 NAEP Writing Achievement Levels-Setting Project

Dr. Bill Auty

Consultant (Former Assistant Superintendent, Oregon Department of Education)

Dr. Wayne Camara

Executive Vice President, Research and Development, the College Board

Dr. Barbara Dodd

Professor, University of Texas–Austin

Dr. Matthew Johnson

Associate Professor of Statistics and Education, Teachers College, Columbia University

Dr. Mary Pitoniak

Strategic Advisor for Statistical Analysis, Data Analysis, and Psychometric Research,
ETS representative

Dr. Mark Reckase

University Distinguished Professor, Michigan State University

Appendix B

Dear [first and last name]:

The National Assessment Governing Board, an independent part of the U.S. Department of Education, is conducting research that will enable the National Assessment of Educational Progress (NAEP)—known as the “Nation’s Report Card”—to report on the writing skills of students in grades 8 and 12 in the United States based on the results of the first computer-based assessment in the 42-year history of NAEP. The contract to develop achievement levels on the NAEP for writing grades 8 and 12 in 2011 and grade 4 in 2013 was awarded to Measured Progress in Dover, NH, on September 23, 2010. Measured Progress, a nonprofit organization offering customized assessment products and educational services, is conducting a series of standard-setting studies to determine the achievement levels of the eighth-grade and twelfth-grade students who participate in the 2011 NAEP for writing. The first study is a field trial focusing on the 2011 NAEP for writing for grade 12 and will be held in southern New Hampshire.

Measured Progress is currently recruiting educators of students in grade 8 or above in southern New Hampshire and Maine to participate in the field trial and would greatly appreciate your assistance in nominating well-qualified candidates. All qualified panelists must be familiar with the writing skills expected of students in grade 12. Each candidate must have:

- five years of teaching experience;
- two years of experience teaching writing, composition, or journalism; and
- recognition as “outstanding” by someone in a position to make that evaluation.

We are seeking teachers currently active in classrooms for grades 8 and above. If selected, panelists will be expected to do the following types of tasks:

- Prepare for the study by reading advance materials offered online.
- Work with other panelists in small groups and large sessions to classify students’ responses to writing prompts during a two-day standard-setting process to be held July 9–10, 2011, in southern New Hampshire.

Panel members need not have prior experience with this kind of study. They will be trained in the standard-setting methodology by expert facilitators. For their participation, panelists will receive a 10.1-inch netbook and will be reimbursed for travel expenses consistent with federal travel requirements.

We are certain you know teachers with the required expertise and would greatly appreciate your nominations. Please submit your nominations at your earliest convenience by accessing the following link: XXXXXXXXXXXXXXXXXXXX
On the nomination page, you will be asked to provide contact information for each nominee and indicate each nominee’s experience. The deadline for nominations is XX/YY/ZZZZ.

We encourage you to nominate yourself, if you are qualified. Please also feel free to forward this email to qualified colleagues. If you have any questions or have difficulty accessing the form, please submit nominations directly to Dr. Luz Bay at 603.749.9102 or naep-als@measuredprogress.org.

Thank you for your assistance in identifying nominees from New Hampshire and Maine for this important study.

Sincerely,

Luz Bay, Ph.D.
Assistant Vice President
Measured Progress
100 Education Way, Dover, NH 03820
603.749.9102

Appendix C

Dear [first and last name]:

The National Assessment Governing Board, an independent part of the U.S. Department of Education, is conducting research that will enable the National Assessment of Educational Progress (NAEP)—known as the “Nation’s Report Card”—to report on the writing skills of students in grades 8 and 12 in the United States based on the results of the first computer-based assessment in the 42-year history of NAEP. The contract to develop achievement levels on the NAEP for writing grades 8 and 12 in 2011 and grade 4 in 2013 was awarded to Measured Progress in Dover, NH, on September 23, 2010. Measured Progress, a nonprofit organization offering customized assessment products and educational services, is conducting a series of standard-setting studies to determine the achievement levels of the eighth-grade and twelfth-grade students who participate in the 2011 NAEP for writing.

Measured Progress is currently recruiting educators of students in grades 8 and higher to participate in the pilot study or the Achievement Levels Setting (ALS) meeting and would greatly appreciate your assistance in nominating well-qualified candidates. All qualified panelists must be familiar with the writing skills expected of students in grade 8 or 12. Each candidate must have:

- five years of teaching experience;
- two years of experience teaching writing, composition, or journalism; and
- recognition as “outstanding” by someone in a position to make that evaluation.

We are seeking teachers currently active in classrooms for grades 8 and above. If selected, panelists will be expected to do the following types of tasks:

- Prepare for the study by reading advance materials offered online.
- Work with other panelists in small groups and large sessions to classify students’ responses to writing prompts during a four-day standard-setting process. The pilot study will be held November 15–18, 2011, in St. Louis, Missouri, and the ALS meeting February 7–10, 2012, also in St. Louis.

Panel members need not have prior experience with this kind of study. They will be trained in the standard-setting methodology by expert facilitators. For their participation, panelists will receive a 10.1-inch netbook and will be reimbursed for travel expenses consistent with federal travel requirements.

We are certain you know teachers with the required expertise and would greatly appreciate your nominations. Please submit your nominations at your earliest convenience by following the link provided below:

XXXXXXXXXXXXXXXXXXXXXX.

On the nomination page, you will be asked to provide contact information for each nominee and indicate each nominee's experience. The deadline for nominations is *XX/YY/ZZZZ*.

We encourage you to nominate yourself, if you are qualified. Please also feel free to forward this email to qualified colleagues. If you have any questions or have difficulty accessing the form, please submit nominations directly to Dr. Luz Bay at 603.749.9102 or naep-als@measuredprogress.org.

Thank you for your assistance in identifying nominees from your district for this important study.

Sincerely,

Luz Bay, Ph.D.
Assistant Vice President
Measured Progress
100 Education Way, Dover, NH 03820
603.749.9102