

# National Assessment Governing Board Committee on Standards, Design and Methodology

Friday, August 4, 2017  
10:10 am – 12:15 pm

## AGENDA

---

10:10 – 10:15 am	Welcome and Review of Agenda  <i>Andrew Ho, COSDAM Chair</i>	
10:15 – 10:50 am	Next Steps for Implementing Strategic Vision (SV #2-10)  <i>Sharyn Rosenberg, Assistant Director for Psychometrics</i>	Attachment A
10:50 – 11:50 am	Potential NAEP Linking Studies (SV #2)  <i>Enis Dogan, NCES</i>  Discussion: Priorities for NAEP Linking Studies to Fulfill Strategic Vision (SV#2)  <i>Andrew Ho</i>	Attachment B
11:50 am – 12:15 pm	Writing Grade 4 Achievement Levels Setting Project Update  <i>Tim O'Neil, Pearson</i>	Attachment C
	Information Item: Technical Support Contract (SV#2-10)	Attachment D

---

### **Next Steps for Implementing Strategic Vision**

During the November 2016 board meeting, a [Strategic Vision](#) was formally adopted to guide the Board's work over the next several years, with a general goal of increasing the impact of NAEP through increased dissemination and innovation. At the March 2017 board meeting, COSDAM discussed a proposed list of draft activities for which the committee was assigned primary responsibility. COSDAM members noted that the proposed activities seemed reasonable but that it would be helpful to better understand how each activity might be implemented.

The Governing Board staff is working on a plan for documenting milestones and timelines for all Strategic Vision activities using project management software. In the meantime, a preliminary list of next steps has been drafted for each of the activities primarily assigned to COSDAM. Please note that many of the Strategic Vision activities require collaboration across committees and with NCES, but the specific opportunities for collaboration are not explicitly referenced in the table below. In addition, the activities that include contributions from COSDAM but are primarily assigned to another committee (e.g., framework update processes) or a task force (i.e., exploring new approaches to postsecondary preparedness) also have not been included below. Finally, the table does not yet specify details about timelines.

During this session, Governing Board Assistant Director for Psychometrics Sharyn Rosenberg will briefly describe the proposed plans for addressing each of COSDAM's Strategic Vision activities.

#### **Discussion Questions**

1. Has COSDAM's work on the Strategic Vision been adequately captured?
2. What activities are the greatest priorities?
3. Are any of the proposed plans not worthwhile to conduct?
4. Are there other activities that COSDAM should be pursuing?

Strategic Vision Activity	Current Status	Planned Next Steps	Desired Outcome
<p><i>2a: Incorporate ongoing linking studies to external measures of current and future achievement in order to evaluate the NAEP scale and add meaning to the NAEP achievement levels in reporting. Consider how additional work could be pursued across multiple subject areas, grades, national and international assessments, and longitudinal outcomes.</i></p>	<p>COSDAM discussion at May 2017 board meeting to examine how existing findings may be used to add meaning to scale scores and achievement levels, and what additional studies to take on</p> <p>Ongoing linking studies include: national NAEP-ACT linking study; longitudinal studies at grade 12 in MA, MI, TN; longitudinal studies at grade 8 in NC, TN; NAEP-TIMSS linking study; NAEP-HSLS linking study; planned studies by NAEP Validity Studies (NVS) panel</p>	<p>Complete ongoing studies</p> <p>Decide what new studies to take on</p> <p>Decide how to use and report existing and future results</p> <p>Complete additional studies</p>	<p>NAEP scale scores and achievement levels may be reported and are better understood in terms of how they relate to other important indicators of interest (i.e., other assessments and milestones)</p>
<p><i>3e: Research when and how NAEP results are currently used (both appropriately and inappropriately) by researchers, think tanks, and local, state and national education leaders, policymakers, business leaders, and others, with the intent to support the appropriate use of NAEP results (COSDAM with R&amp;D and ADC)</i></p> <p><i>3f: Develop a statement of the intended and unintended uses of NAEP data using an anticipated NAEP Validity Studies Panel (NVS) paper and the Governing Board's research as a resource (COSDAM with NCES).</i></p>	<p>Ina Mullis of the NVS panel spoke with COSDAM at the March 2017 board meeting and is working on a white paper about appropriate uses of NAEP</p> <p>Procurement for Technical Support contract specifies that the research study topic for year 1 will focus on how NAEP results are used by various stakeholders</p>	<p>Use research to draft short document of intended and appropriate uses for Board discussion (November 2018)</p> <p>NCES produces documentation of validity evidence for intended uses of NAEP scale scores</p> <p>Governing Board produces documentation of validity evidence for intended uses of NAEP achievement levels</p>	<p>Board adopts formal statement or policy about intended uses of NAEP. The goal is to increase appropriate uses and decrease inappropriate uses (in conjunction with dissemination activities to promote awareness of this document).</p>

<b>Strategic Vision Activity</b>	<b>Current Status</b>	<b>Planned Next Steps</b>	<b>Desired Outcome</b>
<p><i>5c: Consider new approaches to creating and updating the achievement level descriptors and update the Board policy on achievement levels.</i></p>	<p>Initial conversations conducted with 7 standard setting experts in March/April 2017</p> <p>COSDAM discussion at May 2017 board meeting about scope and process of revising the achievement levels policy</p>	<p>Conduct literature review/synthesis of best practices for creating and updating achievement level descriptors (ALDs)</p> <p>Convene expert panel to discuss technical issues and recommendations for achievement levels policy, including specific guidance about ALDs</p> <p>Draft revised policy statement for Board discussion</p> <p>Seek external feedback and public comment</p> <p>Revised policy statement for Board discussion and ultimately adoption</p>	<p>Board has updated policy on achievement levels that meets current best practices in standard setting and is useful for guiding the Board's achievement levels setting work.</p>
<p><i>7a: Support development and publication of multiple papers exploring policy and technical issues related to NAEP Long-Term Trend. In addition to the papers, support symposia to engage researchers and policymakers to provide stakeholder input into the Board's recommendation.</i></p>	<p>White papers commissioned and posted to Governing Board website (February 2017), symposium held in Washington, DC (March 2017), and follow-up event held at American Educational Research Association (AERA) annual conference (April 2017)</p>	<p>Ongoing board discussion about options for the future of LTT and what additional information may be needed</p>	<p>Determine whether changes to the NAEP LTT schedule are needed (7b) and/or whether changes to the design and administration of the LTT assessment are needed (7c)</p>

Strategic Vision Activity	Current Status	Planned Next Steps	Desired Outcome
<i>9b: Pending outcomes of stakeholder input (9a), evaluate the technical implications of combining assessments, including the impact on scaling and trends.</i>	This activity is entirely dependent on activity 9a (ADC’s determination that it is advisable to combine multiple subject area frameworks from a content perspective) and on the particular subjects that may be combined.	TBD	TBD
<i>10a: Continue research to gather validity evidence for using 12<sup>th</sup> grade NAEP reading and math results to estimate the percentage of grade 12 students academically prepared for college.</i>	Several studies are ongoing (see activity 2a)	<p>Decide whether additional research should be pursued at grade 8 to learn more about the percentage of students “on track” to being academically prepared for college by the end of high school</p> <p>Decide whether Board should make stronger statement and/or set “benchmarks” rather than current approach of “plausible estimates”</p> <p>Decide whether additional research should be conducted with more recent administrations of NAEP and other tests.</p>	Statements about using NAEP as an indicator of academic preparedness for college continue to be defensible and to have appropriate validity evidence.



## Potential Linking Studies for the National Assessment of Educational Progress

The National Center for Education Statistics (NCES) and the National Assessment Governing Board have conducted linking studies to provide useful inferences and information relevant to educational policy which require data that NAEP does not routinely collect. By linking NAEP to another assessment or NCES survey, new relationships with assessment scores, contextual information, and/or outcome variables can be established. For example, in 2011 there was a study that linked the NAEP scale to the Trends in International Mathematics and Science Study (TIMSS) scale so that states could compare the performance of their students with that of students in other countries, without taking TIMSS.

The purpose of this document is to identify linking study opportunities by examining administration schedules of various surveys, including assessment surveys, and potential topics based on the identified opportunities. Identified opportunities will be useful in preparing well-designed linking study designs. Of course, execution of any potential linking study is contingent on the availability of funds and the availability of data from non-NAEP surveys or assessments.

All NCES survey programs' administration schedules were examined and tabulated to first identify those surveys that have overlap administrations with NAEP in the near future and then examine the specific features such as target population of overlapped surveys and indicator variables and/or report SES measures. More than 40 NCES survey programs were examined, and 22 programs were identified to have at least one overlap administration. Not only overlap in future administrations, overlap in recent administrations (from 2013 to 2016) also were examined, as there should be some useful linking study opportunities using already collected data.

For each survey program, in addition to information from websites and reports, technical documentation, whenever available, was reviewed to collect any relevant information. A few examples of the major types of information collected are as follows:

- target populations and respondents (if different)
- mode of administration
- variables measured

Table 1 below lists potential linking studies. Included here are some ongoing studies including the Early Childhood Longitudinal Study with NAEP Reading data collected in 2015, and the Trends in International Mathematics and Science Study with NAEP Mathematics and Science collected in 2015. Table 2 lists the level of data collected for each survey.

**Table 1: NCES surveys that overlap with NAEP**

	Year: 20__											
Survey	13	14	15	16	17	18	19	20	21	22	23	24
NAEP	Reading Math [4/8/12]	History, Civics, Geography TEL [8]	Reading, Math Science [4/8/12]	Arts [8]	Reading Math Writing [4/8]	History, Civics, Geography TEL [8]	Reading, Math Science [4/8/12]	†	Reading Math Writing [4/8/12]	History, Civics, Geography TEL [8]	Reading, Math Science [4/8/12]	Arts [8] For Lang [G12]; LTT
Common Core of Data F-33	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Private School Universe Survey							✓		✓		✓	
School Survey on Crime and Safety	✓		✓		✓		✓		✓		✓	
National Teacher & Principal Survey					✓		✓		✓		✓	
OECD International Early Learning Study											✓	
The Early Childhood Longitudinal Study Reading collected in 2015			✓ gr. 4							✓		
Progress in International Reading Literacy Study				✓					✓			
Trends in International Mathematics and Science Study			✓				✓				✓	
Middle Grades Longitudinal Study								✓				
The International Computer and Information Literacy Study					✓							
Program for the International Assessment of Adult Competencies										✓		
High School Longitudinal Study Linked to Gr 12 Math in 2012	✓ (first year of college)			✓ (senior college)								
High School and Beyond 2020												✓Gr. 12

† Not applicable. There are no assessments listed for 2020 on the Governing Board schedule.

**Table 2: Level of data collected by other NCES surveys**

CCD	School, District, State level
ECLS-K	Student, parents, teachers, school admins
F-33	District level (local education agencies or LEAs)
HSLs	Student, school admins, teacher, parents
High School and Beyond	Student, school admins, teacher, parents
MGLS	Student level, grades 6-8
NTPS-Teacher Principal Survey	Teachers/principals. Formerly the SASS
PIAAC	Adults, international
PIRLS	Students, teachers, schools, international
PSS-Private School Universe Survey	School level
SCS – School Crime Supplement to NCVS	Student level
TIMSS	Student, teacher, school level, international



## Developing Achievement Levels for the National Assessment of Educational Progress Writing at Grade 4



**Purpose:** The purpose of this session is to provide an update to the Committee on Standards, Design and Methodology (COSDAM) regarding the development of achievement levels for the 2017 NAEP Grade 4 Writing. *In this session, Tim O'Neil, NAEP Grade 4 Writing Achievement Levels-Setting (ALS) Project Director for Pearson, will provide a brief update on the project.*

Legend:  
 Light shading: Completed  
 No shading: To be completed after 8/04/17

**Project Overview:** On August 3, 2016, the National Assessment Governing Board (Governing Board) awarded a contract to Pearson (as a result of a competitive bidding process) for developing achievement levels for the National Assessment of Educational Progress (NAEP) for grade 4 writing. The 2017 Grade 4 NAEP Writing assessment is the first administration of the grade 4 assessment developed to meet the design specifications described in the current computer-based Writing Framework. The assessment is a digital-based assessment, comprised of constructed response items, for which students compose and construct their responses using word processing software on a tablet. The assessment was administered to a nationally representative sample of approximately 22,000 grade 4 students in the spring of 2017.<sup>1</sup>

Dr. Tim O’Neil is the grade 4 writing ALS project director at Pearson and Dr. Marc Johnson is the assistant project director at Pearson. Pearson will conduct a field trial, a pilot study, and an achievement levels-setting (ALS) meeting and produce a set of recommendations for the Governing Board to consider in establishing achievement levels for the grade 4 NAEP writing assessment. The Governing Board is expected to take action on the writing grade 4 achievement levels during the May 2018 meeting. Pearson will utilize a body of work (BoW) methodology using Moodle software to collect panelist ratings and present feedback. Dr. Lori Nebelsick-Gullet will serve as the process facilitator for the pilot and operational ALS meetings; Victoria Young will serve as the content facilitator for the pilot and operational ALS meetings; and Drs. Susan Cooper Loomis and Steven Fitzpatrick will serve as consultants.

For setting standards, Pearson will use a body of work approach in which panelists will make content-based cut score recommendations. The body of work methodology is a holistic standard setting method for which panelists evaluate sets of examinee work (i.e., bodies of work) and provide a holistic judgment about each student set. These content-based judgments will be made over three rounds. The process to be implemented for the standard setting meeting follows body of work procedures used in previous NAEP standard setting studies. In addition, a field trial will be conducted prior to the pilot study which will provide an opportunity to try out a number of key aspects of the ALS plan, including the logistical design of the ALS studies such as the use of tablets and laptop computers, the ease with which the panelists can enter judgments and questionnaire responses, and the arrangement of tables and panelists.

The Governing Board policy on Developing Student Performance Levels for NAEP (<https://www.nagb.org/content/nagb/assets/documents/policies/developing-student-performance.pdf>) requires appointment of a committee of technical advisors who have expertise in standard setting and psychometrics in general, as well as issues specific to NAEP. These advisors will be convened for 8 in-person meetings and up to 6 webinars to provide advice at every key point in the process. They provide feedback on plans and materials before activities are implemented and review results of the process and analyses. Six external experts in standard setting are serving on the Technical Advisory Committee on Standard Setting (TACSS):

---

<sup>1</sup> Achievement levels were set for Writing grades 8 and 12 with the 2011 administration of those assessments. The grade 4 assessment initially was scheduled to be administered in 2013 but the Governing Board postponed it to 2017 due to budgetary constraints.

**Dr. Gregory Cizek**

Professor of Educational Measurement, University of North Carolina at Chapel Hill

**Dr. Barbara Dodd**

Professor of Professor of Quantitative Methods, University of Texas at Austin

**Dr. Steve Ferrara**

Independent Consultant

**Dr. Matthew Johnson**

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

**Dr. Vaughn G. Rhudy**

Executive Director, Office of Assessment, West Virginia Department of Education

**Dr. Mary Pitoniak**

Senior Strategic Advisor for Statistical Analysis, Data Analysis, and Psychometric Research, Educational Testing Service (NAEP Design, Analysis, and Reporting Contractor)

**August 2017 Update:**

*Field Trial Conducted June 5<sup>th</sup> and 6<sup>th</sup> in San Antonio, TX*

**Purpose:** The purpose of the 2-day field trial was to test aspects of the logistical design of the ALS procedures such as panelists' use of the NAEP tablets and laptop computers, and the ease with which panelists use the Moodle software platform to enter judgments, respond to evaluations, and evaluate feedback. The meeting also allowed us to gauge the effectiveness of training on most of the activities to be carried out during the pilot and operational ALS meetings (each planned for 3.5 days). Given the abbreviated schedule which allowed for only one round of judgments, along with some other important differences between the field trial and the upcoming pilot and operational meetings, the field trial was not intended to produce ALS results for the Board's consideration.

Panelists were recruited regionally (as opposed to nationally, as will be for the pilot and operational meetings). Twenty of the most highly qualified nominees were chosen as panelists. Of these, 55 percent were teachers currently engaged in writing instruction at the grade 4 level; 15 percent were non-teacher educators; and 30 percent general public – all of whom had writing expertise and were experienced with children in the fourth grade. One panelist withdrew at the last minute. Due to proximity to the beginning of the meeting and since pre-meeting activities with the remaining panelists had already begun, the meeting proceeded with the remaining 19 participants.

**Pre-Meeting Activities:** A welcome notification was sent to all participants with a packet of printed pre-meeting materials two weeks prior to the meeting (cover letter, agenda, briefing booklet, policy and subject-specific ALDs, Moodle users' guide, and NAEP Writing Framework). The Moodle site went live May 24th (1.5 weeks before the meeting) and panelists

were requested to log into the site, review several materials, and complete a pre-meeting process evaluation. By the start of the meeting, all panelists had logged in and worked with the Moodle site.

Overall, evaluation responses suggested most panelists came away reasonably well informed after reviewing the briefing booklet, NAEP Writing Framework, and ALDs and that Moodle was generally effective for their review (see Appendix – Pre-Meeting).

**Day 1 Activities:** The meeting began with panelists arriving, signing in, and finding their pre-assigned seats. The meeting was fully supported on site by the following staff who oversaw all aspects of the meeting, including technical support of the NAEP tablets:

National Assessment Governing Board Staff:

- Sharyn Rosenberg, Ph.D., Assistant Director for Psychometrics and Contracting Officer's Representative

Pearson Staff:

- Tim O'Neil, Ed.D., Project Director
- Marc Johnson, Ph.D., Assistant Project Director
- Trey Heideman, Research Assistant
- John Hanson, Program Manager
- Julie Downey, Senior Project Manager
- Eric Moyer, Ph.D., Moodle Architect and on-site support
- Lori Nebelsick-Gullett, Ph.D., Process facilitator
- Victoria Young, Content facilitator

Observers:

- Amy Yamashiro, Ph.D., National Center for Educational Statistics
- Anderson Davis, Fulcrum
- Dan Weber, Westat

The meeting was conducted according to the (attached) agenda. Overall evaluation results for the early activities suggested training and orientation was clear and appropriately timed (see Appendix – Evaluation 1). The exception to this was training on reviewing and applying scoring rubrics to panelist responses, where roughly a quarter of panelists thought more time should be devoted. This was partly the result of having to work around a NAEP tablet issue where panelist responses were not available for many panelists. During the afternoon activities, the timing allotted to the classification task allowed for roughly an hour to classify 18 responses (one per rubric score across 3 writing purposes) with some room for discussion. The activity proceeded without issue other than there was too little time to complete the activity. The day closed promptly at 5:00 pm and panelists were dismissed.

**Day 2 Activities:** Since timing was an issue at the end of Day 1 with the ALD classification judgment task, we sought to be more efficient in the training activities around the BoW judgment task to build in more time for the actual task (planned to be a 2 hour task). Given that there had already been an introduction to the BoW methodology and reference to the judgment activity

involving two responses each, this was felt to be feasible without hindering panelists' understanding of their task. We were able to work through these activities and leave panelists with an extra hour of time for the judgment task. The judgment task was started before lunch. During lunch, panelists were instructed not to discuss anything about the judgment task with their peers, as the task is an independent activity. Panelists resumed after lunch. Some were able to judge all 50 BoWs within the three and a quarter hour allotment. But there were several others who took an additional fifteen minutes to complete their ratings beyond the target finish time.

Data from the round of judgments were collected, analyzed, QCed, and feedback tables and graphs posted back to panelists in Moodle. The process took longer than planned for and resulted in having to abbreviate the feedback training and discussion somewhat to accommodate. Feedback tables and charts were provided with training and an abbreviated review and discussion activity was conducted. Final wrap up and discussion about the overall ALS process was held and comments were overall positive. Several panelists thanked us for the opportunity to participate in these meetings. Comments shared were reflected in final process evaluations. Panelists were thanked for their participation and the meeting was adjourned.

**TACSS Meeting on Field Trial:** On June 22<sup>nd</sup>, the TACSS met to review and discuss outcomes from the field trial meeting, including all of the panelist evaluations (see Appendix). The TACSS noted that the field trial was a successful dry run of the most critical elements of the ALS design. The TACSS agreed that panelist evaluations taken prior to and throughout the meeting were positive overall and indicative of effective development and administration of the ALS tools, materials, and procedures. They reviewed and contributed to the following table of lessons learned and associated action plans. In addition, they reviewed and discussed a draft agenda for the pilot study (attached) in light of what was learned from the field trial.

#### **NAEP Grade 4 Writing ALS Field Trial Lessons Learned and Plans of Action**

<b>Meeting Segment</b>	<b>Lesson Learned</b>	<b>Plan of Action</b>
<b>Pre-Meeting</b>	A few panelists were confused about the overall goal of the meeting and what the panelist role entails.	Review and revise the initial cover letter, initial description within Moodle, and front matter within the briefing booklet to better explain the purpose of the pilot and operational meetings as well as the role panelists have in the ALS process.
<b>Day 1</b>	When panelists took a sample assessment on the NAEP tablets, the audio played on speakers even when panelists had earphones on.	NCES and their contractors will modify and review the NAEP tablets accordingly to prevent this occurrence at the upcoming meetings.

<b>Meeting Segment</b>	<b>Lesson Learned</b>	<b>Plan of Action</b>
	<p>When panelists took a sample assessment on the NAEP tablets, several panelist responses were not saved and it was necessary to have panelists review neighbor panelist responses to work through the application of the rubric to the responses. Panelists felt more time should be devoted to this activity.</p>	<p>NCES and their contractors will modify and review the NAEP tablets to ensure panelist responses can be viewed after taking the test. The time dedicated to this activity will be increased for the pilot and operational meetings.</p>
	<p>Logging into the tablets for each use (test, response, prompts) involved several steps and was somewhat burdensome.</p>	<p>Include all login information and passwords on the PowerPoint presentation to help facilitate logging in.</p>
	<p>Panelists had insufficient time for the ALD classification task. It was difficult for panelists to apply knowledge when rushed. Panelists did not have enough time to review all prompts.</p>	<p>More time will be built into the pilot and operational agenda for these tasks.</p>
	<p>Panelists requested to be able to review student responses during the discussion segment of the ALD classification activity.</p>	<p>Will create a student response database in Moodle to allow panelists to refer back to responses during discussion (we currently have this for the judgment task discussions and were able to create one for the practice judgment task on site).</p>
	<p>Several of the non-educator panelists indicated that they had difficulty conceptualizing grade 4 student writing performance without having worked with grade 4 student performance prior to this meeting.</p>	<p>Integrate exemplar student responses into the training activities focused on gauging student performance within the context of the ALDs. Also add exemplar BoWs into the training activities around the judgment task.</p>

<b>Meeting Segment</b>	<b>Lesson Learned</b>	<b>Plan of Action</b>
<b>Day 2</b>	The orientation and training of the BoW classification task was able to be covered in less time than what was planned.	Review time allotments for orientation and training planned for the pilot and operational meetings in relation to the adjacent activities and reduce if warranted.
	Not all panelists understood that the BoWs were ordered from highest to lowest scoring and/or knew what to do with the information.	Per TACSS recommendation, a graphic will be included within the training slides that helps reinforce the ordering of BoWs and presents several example judgment patterns. Patterns would not reflect perfect transitions across achievement levels, but would rather reflect general transitions (varying in degree of noisiness).
	The time allotted for completing the judgement task was too short.	More time will be built into the Pilot and Operational agenda for this task. The first and third rounds of judgments will be arranged to occur at the end of the second and third days to allow panelists additional time if needed.
	Some panelists requested access to a timer or more regular time updates (e.g. a timer displayed on the presentation slides).	A slide will be included within the pilot and operational meeting presentations that incorporates a timer. Also, the process facilitator will make regular periodic updates to panelists with general points of reference to time remaining.
	Some panelists noted distractions/peripheral noise and movement during the activity.	Additional effort will be made during all focused independent activities to maintain quiet.
	The agenda allotted insufficient time to collect, analyze, and QC judgment data, and return feedback to panelist folders.	More time has been built into the pilot and operational agendas for these activities. The first and third rounds of judgments will occur at the end of days.

There were two additional takeaways. One was the discovery that the writing ALDs printed in the 2017 NAEP Writing Framework and used in the field trial were not the correct versions. The corrected versions have since been provided to Pearson and will be incorporated into all relevant activities of the pilot and operational meetings. For reference, the final grade 4 writing ALDs are attached here and are formatted as they will be used within the context of the ALS meetings. These combine both the policy definitions with the writing content descriptions for grade 4.

The second item is that some panelists expressed concern about whether all 4th graders can write on computer, in particular given the time limit on the NAEP assessment (30 minutes per response). For the pilot and operational meetings, the Governing Board presentation will emphasize that NAEP is forward-looking and its writing framework defines the intended construct as “writing on computer”. This presentation will specify that the purpose of the NAEP writing assessment is to obtain a national picture of the extent to which students are able to write on computer. Emphasis will also be made about the fact that there are no individual student- or school-level scores on NAEP (or even state-level scores at this time), and students who may not have had the opportunity to learn how to write on computer will not be penalized in any way.





### NAEP Grade 4 Writing ALS Agenda – Field Trial Study

<b>Day 1</b>		
<b>Time</b>	<b>Agenda Item</b>	<b>Lead</b>
8:00 – 8:30 a.m.	Registration Outside Meeting Room	Pearson
8:30 – 8:40 a.m.	Welcome and Introductions of the Project Team and Observers  Review of Agenda  Overview of Panelist Recruitment	Dr. Tim O’Neil
8:40 – 9:00 a.m.	Information about the National Assessment of Educational Progress (NAEP) and the National Assessment Governing Board	Dr. Sharyn Rosenberg
9:00 – 9:45 a.m.	Overview of: <ul style="list-style-type: none"> <li>● NAEP Grade 4 Writing Achievement Levels-Setting (ALS) Process</li> <li>● Body of Work Methodology</li> </ul>	Dr. Tim O’Neil Dr. Lori Nebelsick-Gullett
9:45 – 10:00 a.m.	Break	
10:00 – 10:30 a.m.	Training on Moodle  Discussion	Dr. Eric Moyer Dr. Lori Nebelsick-Gullett
10:30 – 12:00 p.m.	Take the NAEP Writing Assessment	Dr. Lori Nebelsick-Gullett
12:00 – 12:45 p.m.	Lunch	



### NAEP Grade 4 Writing ALS Agenda – Field Trial Study

<b>Day 1 (Cont.)</b>		
<b>Time</b>	<b>Agenda Item</b>	<b>Lead</b>
12:45 – 1:15 p.m.	Training on Writing Rubric	Dr. Lori Nebelsick-Gullett Victoria Young
1:15 – 2:30 p.m.	Scoring and Discussion of NAEP Writing Performance  Review of NAEP Writing Tasks	Dr. Lori Nebelsick-Gullett Victoria Young
2:30 - 2:45 p.m.	Break	
2:45 – 3:15 p.m.	Understanding the NAEP Writing Framework	Victoria Young
3:15 – 4:45 p.m.	NAEP Writing Achievement Level Descriptors  ALD Practice activity	Victoria Young
4:45 – 5:00 p.m.	End of Day 1 Activities <ul style="list-style-type: none"> <li>● Evaluation 1</li> <li>● Check in Materials</li> <li>● Review of Day 2 Schedule</li> </ul>	Dr. Lori Nebelsick-Gullett
<b>Adjourn</b>		



**NAEP Grade 4 Writing ALS Agenda – Field Trial Study**

<b>Day 2</b>		
<b>Time</b>	<b>Agenda Item</b>	<b>Lead</b>
8:00 – 8:15 a.m.	Reconvene, Recap Day 1, and Review Agenda for Day 2	Dr. Lori Nebelsick-Gullett
8:15 – 9:30 a.m.	Orientation to Body of Work Methodology  Training on BOW Classification	
9:30 - 9:45 a.m.	Break	
9:45 - 10:45 a.m.	Practice Exercise	
10:45 – 11:30 a.m.	Discuss Practice Exercise	
11:30 - 12:15 p.m.	Lunch	
12:15 – 2:15 p.m.	Round of Classifications	Dr. Lori Nebelsick-Gullett
2:15 - 2:30 p.m.	Break	
2:30 – 4:00 p.m.	Classification Feedback <ul style="list-style-type: none"> <li>● individual cut score information</li> <li>● cut score summary table</li> <li>● cut score distribution charts</li> <li>● consequences data</li> </ul> Review of Disparate BoWs Discussion of ALS Process	Dr. Lori Nebelsick-Gullett
4:00 – 4:30 p.m.	Exit Survey/Process Evaluation	



### NAEP Grade 4 Writing ALS Agenda – Field Trial Study

<b>Day 2 (Cont.)</b>		
<b>Time</b>	<b>Agenda Item</b>	<b>Lead</b>
4:30 – 4:45 p.m.	Check in Materials	Pearson
4:45 – 5:00 p.m.	Reimbursement request/ Procedures Wrap Up	Pearson/ Dr. Sharyn Rosenberg
<b>Adjourn</b>		



### NAEP Grade 4 Writing ALS Agenda – Pilot Study (Draft)

<b>Day 1</b>		
<b>Time</b>	<b>Agenda Item</b>	<b>Lead</b>
8:00 – 8:30 a.m.	Registration Outside Meeting Room	Pearson
8:30 – 8:40 a.m.	Welcome and Introductions of the Project Team and Observers  Review of Agenda  Overview of Panelist Recruitment	Dr. Tim O’Neil
8:40 – 9:00 a.m.	Information about the National Assessment of Educational Progress (NAEP) and the National Assessment Governing Board	Dr. Sharyn Rosenberg
9:00 – 9:45 a.m.	Overview of: <ul style="list-style-type: none"> <li>● NAEP Grade 4 Writing Achievement Levels-Setting (ALS) Process</li> <li>● Body of Work Methodology</li> </ul>	Dr. Tim O’Neil Dr. Lori Nebelsick-Gullett
9:45 – 10:00 a.m.	Break	
10:00 – 10:30 a.m.	Training on Moodle  Discussion	Dr. Eric Moyer Dr. Lori Nebelsick-Gullett
10:30 – 12:00 p.m.	Take the NAEP Writing Assessment	Dr. Lori Nebelsick-Gullett
12:00 – 12:45 p.m.	Lunch	



### NAEP Grade 4 Writing ALS Agenda – Pilot Study (Draft)

<b>Day 1 (Cont.)</b>		
<b>Time</b>	<b>Agenda Item</b>	<b>Lead</b>
12:45 – 1:15 p.m.	Training on Writing Rubric	Dr. Lori Nebelsick-Gullett Victoria Young
1:15 – 2:00 p.m.	Scoring and Discussion of NAEP Writing Performance  Review of NAEP Writing Tasks	Dr. Lori Nebelsick-Gullett Victoria Young
2:00 - 2:15 p.m.	Break	
2:15 – 3:00 p.m.	Understanding the NAEP Writing Framework	Victoria Young
3:00 – 4:45 p.m.	NAEP Writing Achievement Level Descriptors  ALD Practice activity	Victoria Young
4:45 – 5:00 p.m.	End of Day 1 Activities <ul style="list-style-type: none"> <li>● Check in Materials</li> <li>● Review of Day 2 Schedule</li> <li>● Evaluation 1a</li> </ul>	Dr. Lori Nebelsick-Gullett
<b>Adjourn</b>		



### NAEP Grade 4 Writing ALS Agenda – Pilot Study (Draft)

Day 2		
Time	Agenda Item	Lead
8:00 – 8:15 a.m.	Reconvene, Recap Day 1, and Review Agenda for Day 2	Dr. Lori Nebelsick-Gullett
8:15 – 9:15 a.m.	ALD Practice activity (Cont.) Discussion Evaluation 1b	Victoria Young
9:15 – 10:00 a.m.	Orientation to Body of Work Methodology  Training on BOW Classification	Dr. Lori Nebelsick-Gullett
10:00 – 10:15 a.m.	Break	
10:15 - 11:15 a.m.	Practice Exercise	Dr. Lori Nebelsick-Gullett
11:15 – 12:15 p.m.	Discuss Practice Exercise	
12:15 - 1:00 p.m.	Lunch	
1:00– 5:00 p.m.	Round 1 Classifications	Dr. Lori Nebelsick-Gullett
5:00- 5:15 p.m.	End of Day 2 Activities <ul style="list-style-type: none"> <li>• Check in Materials</li> <li>• Review Day 3 Schedule</li> </ul>	
Adjourn		



### NAEP Grade 4 Writing ALS Agenda – Pilot Study (Draft)

<b>Day 3</b>		
Time	Agenda Item	Lead
8:00 – 8:15 a.m.	Reconvene, Recap Day 2, and Review Agenda for Day 3	Dr. Lori Nebelsick-Gullett
8:15 – 10:00 a.m.	Round One Feedback Review of Disparate BoWs Evaluation 2	Dr. Lori Nebelsick-Gullett
10:00- 10:15 a.m.	Break	
10:15 – 12:00 p.m.	Round Two Classifications	Dr. Lori Nebelsick-Gullett
12:00- 12:45 p.m.	Lunch	
12:45- 1:45 p.m.	Round Two Feedback Review and Discuss Consequences Data Evaluation 3	Dr. Lori Nebelsick-Gullett
1:45- 2:00 p.m.	Break	
2:00- 5:00 p.m.	Round Three Classifications	Dr. Lori Nebelsick-Gullett
5:00- 5:15 p.m.	End of Day 3 Activities <ul style="list-style-type: none"> <li>• Check in Materials</li> <li>• Review Day 4 Schedule</li> </ul>	
<b>Adjourn</b>		





**NAEP Grade 4 Writing ALS Agenda – Pilot Study (Draft)**

<b>Day 4</b>		
<b>Time</b>	<b>Agenda Item</b>	<b>Lead</b>
8:00 – 8:15 a.m.	Reconvene, Recap Day 3, and Review Agenda for Day 4	Dr. Lori Nebelsick-Gullett
8:15 – 10:00 a.m.	Round Three Feedback Review and Discuss Consequences Data Consequences Questionnaire Evaluation 4	
10:00- 10:15 a.m.	Break	
10:15 – 11:15 a.m.	Training on Identification of Exemplar BoWs Identification of Exemplar BoWs	Dr. Lori Nebelsick-Gullett
11:15- 11:45 a.m.	Exit Survey/Process Evaluation	
11:45- 12:00 p.m.	End of Meeting Activities <ul style="list-style-type: none"> <li>• Check in Materials</li> <li>• Reimbursement/Wrap Up</li> </ul>	Pearson Dr. Sharyn Rosenberg
<b>Adjourn</b>		



## Writing Achievement Level Descriptions - Grade 4

NAEP Policy definitions and subject specific Achievement Level Descriptions define *what students should know and be able to do* at three levels: *Basic*, *Proficient*, and *Advanced*.

<b>Basic</b>	Policy Definition: This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.
	Fourth-grade students writing at the <i>Basic</i> level should be able to address the tasks appropriately and at least partially accomplish their communicative purposes. Texts should be appropriately structured. Many of the ideas in the texts should be developed, and their texts should include supporting details and examples that are relevant to the topic, purpose, and audience. Most sentences should be well structured, and texts may be composed mostly of simple sentences. Many of the words and phrases should be appropriate to the topics, purposes, and audiences. Spelling, grammar, usage, capitalization, and punctuation skills should be sufficiently accurate to convey general meaning, although there may be some errors that detract from meaning.
<b>Proficient</b>	Policy Definition: This level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter.
	Fourth-grade students writing at the <i>Proficient</i> level should be able to address the tasks appropriately and accomplish their communicative purposes. Texts should be appropriately structured and coherent. Most of the ideas in their texts should be developed effectively, and their texts should include supporting details and examples that support the main ideas. Texts should have well structured sentences and a variety of sentence types—simple, compound, and complex. Words and phrases should be thoughtfully selected and appropriate to the topics, purposes, and audiences. Spelling, grammar, usage, capitalization, and punctuation should be sufficiently accurate to communicate clearly with the reader. There may be some errors in the texts, but these errors should not impede meaning.
<b>Advanced</b>	Policy Definition: This level signifies superior performance beyond proficient.
	Fourth-grade students writing at the <i>Advanced</i> level should be able to address the tasks appropriately and accomplish their communicative purposes in effective ways. Texts should be well structured and coherent. The ideas in the texts should be developed fully and effectively. Their texts should include supporting details and examples that are closely related to the topic, purpose, and audience and that enhance communicative effectiveness. Sentences should be well structured, and texts should include a variety of sentence types (simple, compound, and complex) to enhance their communicative effectiveness. Words and phrases should be chosen skillfully, and they should both enrich meaning in the texts and enhance communicative effectiveness. Spelling, grammar, usage, capitalization, and punctuation should be mostly accurate and well developed, and they should be used appropriately. Grammatical, mechanical, and usage choices should contribute to communicative effectiveness. There may be a few errors, but they should not impede meaning.

## APPENDIX

### FIELD TRIAL PROCESS EVALUATION RESULTS

## Pre-Meeting

Question	Options						Average	SD
	Strongly disagree (1)	Disagree (2)	Slightly disagree (3)	Slightly agree (4)	Agree (5)	Strongly agree (6)		
The briefing booklet was clear.	2 11%	0 0%	0 0%	0 0%	6 33%	10 56%	5.1	1.5
The Moodle interface was an effective tool for reviewing the briefing booklet.	1 6%	0 0%	2 11%	3 17%	6 33%	6 33%	4.7	1.3
After my review, I have a good general understanding of the NAEP Writing Framework.	2 11%	0 0%	0 0%	1 6%	11 61%	4 22%	4.7	1.4
The Moodle interface was an effective tool for reviewing the NAEP Writing Framework.	2 11%	0 0%	1 6%	4 22%	7 39%	4 22%	4.4	1.5
The NAEP policy definitions were clear.	2 11%	0 0%	1 6%	0 0%	8 44%	7 39%	4.8	1.5
The NAEP grade 4 writing ALDs were clear.	2 11%	0 0%	1 6%	0 0%	10 56%	5 28%	4.7	1.5
The Moodle interface was an effective tool for reviewing the policy definitions and grade 4 writing ALDs.	1 6%	0 0%	3 17%	3 17%	6 33%	5 28%	4.6	1.3

Question	Options					
	Did not review (1)	Less than 15 minutes (2)	15-30 minutes (3)	31-45 minutes (4)	46-60 minutes (5)	More than 60 minutes (6)
How much time did you spend reviewing the briefing booklet?	0 0%	0 0%	1 6%	7 39%	4 22%	6 33%
How much time did you spend reviewing the NAEP Writing Framework?	0 0%	1 6%	1 6%	6 33%	2 11%	8 44%
How much time did you spend reviewing the NAEP policy and subject specific ALDs?	0 0%	4 22%	6 33%	5 28%	1 6%	2 11%

**Prior to showing up at the field trial meeting, is there anything in particular about the NAEP Writing Framework that you need clarified? If none, please write 'none.'**

- None
- Even after reading all the material, the main goal of the participants remains elusive to me.
- none
- none
- None
- None
- none
- none
- None
- None
- none
- None
- None. However, I'm still not clear about what my role is as a participant of this field trial.
- none
- none
- none
- none

**What questions do you have regarding the content of the NAEP policy definitions and grade 4 writing ALDs? If none, please write 'none.'**

- none
- none
- none
- none
- None
- none
- None
- None
- None
- none
- none
- none
- none
- none
- I have not yet explored the web links for more examples, but examples I believe are key to understanding definitions.
- none
- None
- None

**What questions do you have regarding the material in the briefing booklet? If none, please write 'none.'**

- None
- none
- None
- none
- My only question is in regards to English Language Learners. Will students be given an option to have the writing task read to them?
- none
- I felt that you needed some clearer statement about the purpose of this meeting. It seems vague currently. Just something up front.
- none
- None
- None
- none
- none
- none
- None
- none
- None
- Besides your reference materials, do I need to bring anything else?

**Prior to showing up at the field trial, do you have particular questions regarding your role in the Field Trial? If none, please write 'none.'**

- None
- none
- none
- none
- None
- None
- none
- None.
- none
- no specific questions, though to be honest I am still not clear on my role
- none
- None
- none
- none
- none
- None
- If this is the case, I have concerns about writing at the level of a student and having that serve as a standard. As an educator, I take the students writing, as they come to me, and help them improve it.
- I understand that student writing samples are utilized to set achievement level expectations, but I am uncertain about how my writing will fit into this equation. I'm excited to see how this will all come together.
- None

## Process Evaluation 1

Question	Options						Average	SD
	Strongly disagree (1)	Disagree (2)	Slightly disagree (3)	Slightly agree (4)	Agree (5)	Strongly agree (6)		
The advanced materials helped prepare me for my role as a panelist.	0 0%	0 0%	0 0%	3 16%	6 32%	10 53%	5.4	0.7
The explanation of the overall NAEP program was clear.	0 0%	0 0%	0 0%	4 21%	1 5%	14 74%	5.5	0.8
The explanation of the purpose of the NAEP Achievement Level Setting (ALS) meeting was clear.	0 0%	0 0%	1 5%	2 11%	5 26%	11 58%	5.4	0.9
The training provided on how to use the ALS computer to access the Moodle interface was clear.	0 0%	0 0%	0 0%	1 5%	6 32%	12 63%	5.6	0.6
The training provided on how to perform assigned tasks using the Moodle interface was clear.	0 0%	0 0%	0 0%	2 11%	4 21%	13 68%	5.6	0.7
The Moodle interface was an effective tool for reviewing all pre-meeting materials made available prior to the field trial.	0 0%	0 0%	0 0%	6 32%	4 21%	9 47%	5.2	0.9
The training the facilitators provided for using a tablet to take the grade 4 NAEP writing assessment was clear.	0 0%	0 0%	1 5%	0 0%	6 32%	12 63%	5.5	0.8
The training by the facilitators for using a tablet to take the assessment provided the information I needed to complete that task.	0 0%	0 0%	0 0%	2 11%	5 26%	12 63%	5.5	0.7
The training on the NAEP writing scoring rubrics was clear.	0 0%	0 0%	0 0%	5 26%	6 32%	8 42%	5.2	0.8

Question	Options					Average	SD
	Far too little (1)	Slightly too little (2)	About right (3)	Slightly too much (4)	Far too much (5)		
The amount of time spent on the General Orientation to the NAEP program was	0 0%	1 5%	15 79%	3 16%	0 0%	3.1	0.4
The amount of time allocated for the general introduction to the NAEP Achievement Level Setting process was	0 0%	1 5%	17 89%	1 5%	0 0%	3.0	2.4
The amount of time spent on orienting me to use of the Moodle interface was	0 0%	0 0%	17 89%	1 5%	1 5%	3.2	2.3
The amount of time facilitators took to provide training on tablet use was	0 0%	1 5%	15 79%	3 16%	0 0%	3.1	2.3
The amount of time provided for reviewing and applying the NAEP writing scoring rubrics was	0 0%	5 26%	11 58%	3 16%	0 0%	2.9	2.6
The amount of time provided for reviewing the NAEP grade 4 writing tasks was	0 0%	0 0%	19 100%	0 0%	0 0%	3.0	2.4



## Process Evaluation 2

Question	Options						Average	SD
	Strongly disagree (1)	Disagree (2)	Slightly disagree (3)	Slightly agree (4)	Agree (5)	Strongly agree (6)		
I am comfortable using the ALS laptop to access the Moodle interface.	0 0%	0 0%	0 0%	0 0%	5 26%	14 74%	5.7	0.4
The presentation of the Writing Framework was clear.	0 0%	0 0%	1 5%	2 11%	8 42%	8 42%	5.2	0.8
The concept of writing at grade 4 as defined by the NAEP Framework is clear to me.	0 0%	0 0%	2 11%	0 0%	11 58%	6 32%	5.1	0.9
The ALDs are clear descriptions about what students should know and be able to do at each achievement level.	0 0%	0 0%	2 11%	3 16%	6 32%	8 42%	5.1	1.0
Reviewing and classifying student responses into ALD categories was helpful in understanding the difference between the ALD categories and the scoring rubrics.	0 0%	1 5%	1 5%	2 11%	9 47%	6 32%	4.9	1.0
I am confident that the panel shares a common understanding of the ALDs.	0 0%	0 0%	2 11%	4 21%	9 47%	4 21%	4.8	0.9
I am confident the panel understands the difference between the ALD categories and the scoring rubrics.	0 0%	0 0%	2 11%	5 26%	10 53%	2 11%	4.6	0.8
The performance (knowledge, skills, and abilities) required to meet the Basic achievement level description is clear to me.	0 0%	0 0%	0 0%	3 16%	9 47%	7 37%	5.2	0.7
The performance (knowledge, skills, and abilities) required to meet the Proficient achievement level description is clear to me.	0 0%	0 0%	0 0%	4 21%	11 58%	4 21%	5.0	0.6
The performance (knowledge, skills, and abilities) required to meet the Advanced achievement level description is clear to me.	0 0%	0 0%	0 0%	2 11%	9 47%	8 42%	5.3	0.7

Question	Options					Average	SD
	Far too little (1)	Slightly too little (2)	About right (3)	Slightly too much (4)	Far too much (5)		
The amount of time allocated for the Writing Framework presentation was	0 0%	0 0%	14 74%	4 21%	1 5%	3.3	0.6
The amount of detail provided during the Writing Framework presentation was	0 0%	0 0%	17 89%	1 5%	1 5%	3.2	0.5
The amount of time allocated for reviewing and discussing the ALDs was	0 0%	6 32%	12 63%	1 5%	0 0%	2.7	0.5
The amount of time allocated for classifying and discussing student responses was	0 0%	12 63%	7 37%	0 0%	0 0%	2.4	0.5

### Process Evaluation 3

Question	Options						Average	SD
	Strongly disagree (1)	Disagree (2)	Slightly disagree (3)	Slightly agree (4)	Agree (5)	Strongly agree (6)		
My understanding of how to use the ALDs to classify a body of work into an achievement level was adequate.	0 0%	0 0%	1 5%	2 11%	9 47%	7 37%	5.2	0.8
The practice training was sufficient for me to classify bodies of work into achievement levels.	0 0%	0 0%	2 11%	0 0%	10 53%	7 37%	5.2	0.9
I understand how student performance on a complete body of work relates to overall performance on the NAEP Writing scale.	0 0%	0 0%	1 5%	3 16%	8 42%	7 37%	5.1	0.9
The Moodle interface was an effective tool for reviewing student bodies of work.	0 0%	0 0%	0 0%	1 5%	8 42%	10 53%	5.5	0.6
The Moodle interface was an effective tool for classifying student bodies of work.	0 0%	0 0%	0 0%	0 0%	9 47%	10 53%	5.5	0.5
I am prepared to classify student bodies of work into achievement levels.	0 0%	0 0%	0 0%	0 0%	10 53%	9 47%	5.5	0.5

  

Question	Options					Average	SD
	Far too little (1)	Slightly too little (2)	About right (3)	Slightly too much (4)	Far too much (5)		
The amount of time allocated for orientation to the achievement level setting process was	0 0%	1 5%	18 95%	0 0%	0 0%	2.9	0.2
The amount of detail provided during the the achievement level setting process presentation was	0 0%	1 5%	17 89%	1 5%	0 0%	3.0	0.3

Please provide an explanation to help us understand what was needed to improve the practice training on the body of work classification process.

- I think it would be a good idea to have samples printed of student work. Then we could read score and identify using a highlighter the specific sentences that swayed our opinion.
- I would like to have an additional set of practice materials.

## Process Evaluation 4

Question	Options						Average	SD
	Strongly disagree (1)	Disagree (2)	Slightly disagree (3)	Slightly agree (4)	Agree (5)	Strongly agree (6)		
The Moodle interface was an effective tool for reviewing student bodies of work.	0 0%	0 0%	0 0%	0 0%	7 37%	12 63%	5.6	0.5
The Moodle interface was an effective tool for classifying student bodies of work.	0 0%	0 0%	0 0%	0 0%	8 42%	11 58%	5.6	0.5
Enter your level of confidence	Average (out of 10) 7.8							
Question	Options					Average	SD	
	Far too little (1)	Slightly too little (2)	About right (3)	Slightly too much (4)	Far too much (5)			
The amount of time allocated for reviewing and classifying 50 bodies of work into achievement levels was	4 21%	7 37%	6 32%	2 11%	0 0%	2.3	0.9	

**Please comment on what was needed to improve the task of reviewing and classifying the student bodies of work.  
If none, please write 'none.'**

- Show the prompt with the essay.
- I needed a little more time.  
Basically, an entire day and a half morning was devoted to redundant explanation of the test. I recommend cutting that in half.
- Break up the reviewing and classifying into three sessions.  
The work space was not so comfortable. I was sitting in a space where my legs were near the legs of the table.
- Each person should have their own table.
- I'm not sure that more training was needed - I simply realize that some of our judgments are subjective and feel a bit of hesitation about certain responses. I try not to overthink it!
- I feel I just needed a bit more time to record adequate notes regarding each judgment I made.
- none  
I would like to see a timer to keep track of how much time I am spending on evaluating each BoW. I began without noticing the time on my computer, which is not correct, so that was my fault, so a timer somewhere would be beneficial so not to spend too much time in the beginning and then not rush at the end to complete.
- I think more practice using a hard copy and discussion among members would be helpful as well as after we give scores, give us the real ones to see if we were correct and the the rationale for them.
- Maybe have an added description for "Below Basic."
- none  
I could have used more time to complete the activity. Also, having the opportunity to take breaks throughout would have helped get through the samples.
- None
- More discussion whole group on various student samples before classifying the 100 BoW
- 50, for me, was too much to score in one sitting. I would have preferred 25 in the AM and 25 in the PM.  
I think I just need more practice, and more back and forth with other panelists (and with you guys) in which we can refer to specific passages and talk about why we rated them as we did. Even talking outside after we all finished, I learned/was persuaded of some things that would change my ratings for some of the borderline Basics. I really felt like I had to SLAM through those responses at top speed, which was a bit exhausting so I was pretty foggy by the end!
- The amount of samples began to run together. Possible break up the exercise into different sections. Then it wouldn't feel so overwhelming.

## Process Evaluation 5

Question	Options						Average	SD
	Strongly disagree (1)	Disagree (2)	Slightly disagree (3)	Slightly agree (4)	Agree (5)	Strongly agree (6)		
I understand how the cut scores were computed for the overall panel.	0 0%	0 0%	1 5%	6 32%	10 53%	2 11%	4.7	0.7
I understand how the cut scores were computed for my subpanel.	0 0%	0 0%	2 11%	4 21%	11 58%	2 11%	4.7	0.8
I understood how to access my Individual Cut Score information using the Moodle interface.	0 0%	0 0%	0 0%	2 11%	9 47%	8 42%	5.3	0.7
I understood how to interpret my Individual Cut Score information and use it during follow-up discussions.	0 0%	0 0%	0 0%	4 21%	13 68%	2 11%	4.9	0.6
<b>Enter your level of confidence</b>	<b>Average (out of 10)</b> 7.16							
I understood how to access the Cut Score Summary Information table using the Moodle interface.	0 0%	0 0%	0 0%	1 5%	8 42%	10 53%	5.5	0.6
I understand how to interpret the information in the Cut Score Summary Information table and use it during follow-up discussions.	0 0%	0 0%	0 0%	4 21%	13 68%	2 11%	4.9	0.6
<b>Enter your level of confidence</b>	<b>Average (out of 10)</b> 7.42							
I understood how to access the Cut Score Distribution charts using the Moodle interface.	0 0%	0 0%	0 0%	2 11%	7 37%	10 53%	5.4	0.7
I understood how to interpret the information in the Cut Score Distribution charts and use it during follow-up discussions.	0 0%	0 0%	0 0%	6 32%	11 58%	2 11%	4.8	0.6
<b>Enter your level of confidence</b>	<b>Average (out of 10)</b> 7.16							

Question	Options						Average	SD
	Strongly disagree (1)	Disagree (2)	Slightly disagree (3)	Slightly agree (4)	Agree (5)	Strongly agree (6)		
I understood how to access the Consequences Data charts using the Moodle interface.	0 0%	0 0%	1 5%	3 16%	7 37%	8 42%	5.2	0.9
I understood how to interpret the information in the Consequences Data charts and use it during follow-up discussions.	0 0%	1 5%	1 5%	3 16%	12 63%	2 11%	4.7	0.9
I found it easy to use the interactive Consequences Data Tool on the Moodle interface.	0 0%	0 0%	1 5%	4 21%	8 42%	6 32%	5.0	0.9
I felt comfortable using the consequences data to evaluate the reasonableness of performance outcomes in relation to	0 0%	0 0%	1 5%	8 42%	9 47%	1 5%	4.5	0.7
Enter your level of confidence	Average (out of 10) 7.16							
The Moodle interface was an effective tool for reviewing and interpreting the feedback information.	0 0%	0 0%	1 5%	0 0%	8 42%	10 53%	5.4	0.7
Question	Options					Average	SD	
	Far too little (1)	Slightly too little (2)	About right (3)	Slightly too much (4)	Far too much (5)			
The amount of detail provided during the presentation on feedback interpretation and use was	0 0%	3 16%	15 79%	1 5%	0 0%	2.9	0.4	
The amount of time allocated for discussing the Cut Score Summary Information table was	0 0%	4 21%	15 79%	0 0%	0 0%	2.8	0.4	
The amount of time allocated for discussing the Cut Score Distribution charts was	0 0%	4 21%	15 79%	0 0%	0 0%	2.8	0.4	
The amount of time allocated for discussing the Consequences Data charts was	1 5%	5 26%	13 68%	0 0%	0 0%	2.6	0.6	

**Please comment on what was needed for easier interpretation and use of the consequences data. If none, please write**

- None
- A bit more time spent on this topic would be helpful.  
If my scores, after putting them into the adjusted cut scores pie would've actually worked. It would've made more sense to me. Or getting to look at data again after round 2 or 3.
- Having more time to use each area and fully understand it.
- none
- a more clear explanation
- I'm not certain at this point exactly what would be needed.
- none
- None
- none
- It would have been helpful to look at example data and discuss its meaning.
- I wanted to talk about my takeaway when it came to the tool. I felt the tool would be helpful if the pie chart did not
- just more time playing around with the info
- Guided process, taking turns how each panelist interpreted each BOW .
- I am really not a statistics oriented individual, so this is a bit out of my grasp. You would need to provide some
- I think just more time, perhaps? Also, maybe slowing down and reminding us more



## Final Evaluation

Question	Options						Average	SD
	Strongly disagree (1)	Disagree (2)	Slightly disagree (3)	Slightly agree (4)	Agree (5)	Strongly agree (6)		
The purpose of this meeting was clear.	0 0%	1 5%	1 5%	2 11%	4 21%	11 58%	5.2	1.2
The advanced materials were helpful.	0 0%	0 0%	0 0%	3 16%	5 26%	11 58%	5.4	0.7
My input was valued by others at my table.	0 0%	0 0%	0 0%	0 0%	5 26%	14 74%	5.7	0.4
My input was valued by others on the overall panel.	0 0%	0 0%	0 0%	1 5%	7 37%	11 58%	5.5	0.6
The facilitation and support provided met my needs as a panelist.	0 0%	0 0%	0 0%	1 5%	2 11%	16 84%	5.8	0.5
The instructions for interpretation and use of the ALDs were clear.	0 0%	0 0%	0 0%	2 11%	5 26%	12 63%	5.5	0.7
The instructions for classifying individual student responses into ALD categories were clear.	0 0%	0 0%	0 0%	1 5%	8 42%	10 53%	5.5	0.6
The instructions for classifying bodies of work into ALD categories were clear.	0 0%	0 0%	0 0%	1 5%	9 47%	9 47%	5.4	0.6
The instructions for interpretation and use of feedback data were clear.	0 0%	0 0%	1 5%	7 37%	4 21%	7 37%	4.9	1.0
During the judgment task, the Achievement Level Descriptions were helpful for placing bodies of work into particular achievement levels.	0 0%	0 0%	0 0%	3 16%	6 32%	10 53%	5.4	0.7

Question	Options						Average	SD
	Strongly disagree (1)	Disagree (2)	Slightly disagree (3)	Slightly agree (4)	Agree (5)	Strongly agree (6)		
The Achievement Level Descriptions were helpful for interpreting and discussing feedback after the round of judgments.	0 0%	0 0%	0 0%	3 16%	3 16%	13 68%	5.5	0.8
Classifying individual student responses into ALD categories during the practice ALD classification activity helped me connect the ALDs to student performance within each of the three writing purposes.	0 0%	0 0%	0 0%	3 16%	6 32%	10 53%	5.4	0.7
Classifying bodies of work into ALD categories during the practice judgment activity helped me connect the knowledge and skills demonstrated through a BoW with a specific achievement level.	0 0%	0 0%	0 0%	3 16%	7 37%	9 47%	5.3	0.7
The instructions the facilitator provided for interpretation and use of the feedback information allowed me to use it effectively during group discussions.	0 0%	0 0%	1 5%	3 16%	8 42%	7 37%	5.1	0.9
After being instructed on the practice ALD classification activity, I understood the task I was to complete for classifying individual student responses into ALD categories.	0 0%	0 0%	0 0%	0 0%	9 47%	10 53%	5.5	0.5
After being instructed on the practice judgment activity, I understood the task I was to complete for classifying bodies of work into ALD categories.	0 0%	0 0%	0 0%	2 11%	6 32%	11 58%	5.5	0.7
Overall, the meeting was well organized.	0 0%	0 0%	0 0%	0 0%	4 21%	15 79%	5.8	0.4
Overall, the meeting afforded the right amount of training to support the stated purposes.	0 0%	0 0%	0 0%	4 21%	7 37%	8 42%	5.2	0.8

Question	Options						Average	SD
	Strongly disagree (1)	Disagree (2)	Slightly disagree (3)	Slightly agree (4)	Agree (5)	Strongly agree (6)		
I was able to distinguish between bodies of work that demonstrated Advanced versus Proficient performance.	0 0%	0 0%	0 0%	6 32%	4 21%	9 47%	5.2	0.9
I was able to distinguish between bodies of work that demonstrated Proficient versus Basic performance.	0 0%	0 0%	1 5%	7 37%	7 37%	4 21%	4.7	0.8
I was able to distinguish between bodies of work that demonstrated Basic versus Below Basic performance.	0 0%	0 0%	0 0%	4 21%	8 42%	7 37%	5.2	0.7
My Individual Cut Score information was helpful during the ALS process.	0 0%	0 0%	2 11%	5 26%	4 21%	8 42%	4.9	1.0
The Cut Score Summary information for my subpanel was helpful during the ALS process.	0 0%	1 5%	1 5%	2 11%	6 32%	9 47%	5.1	1.1
The Cut Score Summary information for the overall panel was helpful during the ALS process.	0 0%	0 0%	1 5%	2 11%	6 32%	10 53%	5.3	0.9
The Consequences Data were helpful for understanding how the cut scores relate to student performance.	0 0%	0 0%	1 5%	4 21%	7 37%	7 37%	5.1	0.9
The discussions of feedback data were helpful during the ALS process.	0 0%	0 0%	0 0%	3 16%	6 32%	10 53%	5.4	0.7
In general, the Moodle interface was helpful for completing the ALS process.	0 0%	0 0%	0 0%	0 0%	7 37%	12 63%	5.6	0.5
Having observers present influenced my judgments.	11 58%	3 16%	2 11%	1 5%	2 11%	0 0%	1.9	1.4

Please indicate any questions or concerns you may have that were NOT answered during this meeting. Include here any suggestions for modifying the Moodle interface or for modifying the process we implemented. If none, please write 'none.'

- Instructions were great, I really understood how to classify BOW's into the correct category, and sway my thinking to other levels.

- None.

- More time for discussion of individual ratings would have been helpful to my own rating process.

I would like more practice in assigning achievement levels, and would like to accomplish this by seeing examples of advanced, proficient, basic, and below basic, along with explanation as to why these BoWs fall into this range. Then I would like to practice it as we did, on the Moodle.

This was an excellent experience for me, and I thank you for allowing me to participate.

- The time constraint places in the body of work activity may have led me to feel rushed and not always make an informed judgement.

- Maybe breaking up the judgement of 50 BoW's into thirds. I would've liked to have had time to write notes and more carefully consider the transition areas.

Judgement of 50 BOW should be divided in different/short sessions. Having breaks in between would be helpful for individuals to stay focused and judge more accurately. Also a little more time for discussion after practice activities can contribute to better understanding of process.

- I'm so grateful for the opportunity to participate in such a real data collecting activity. I feel like a rockstar! Thank you for all your hard work ! LOVED EVERY MINUTE OF THE PROCESS! Thank you!

- none

- none

- none

- None.

- I found having people and moving around distracting at times. Perhaps a carrel would help.

- Tighten Day 1 to include more participation by the panel. Instruction/explanation was too lengthy and often redundant.

1. By the end of Day 1, I was having a hard time listening to the long purely theoretical lecture on the ALDS. It was clear and the speaker was lively; it was just the seventh hour of abstraction when I'd never seen a student response. I had nothing to hang the abstraction onto. I wonder if you could move the practice earlier in the Day 1, and then have Victoria lead all of us in looking over the tricky ones and using the ALDS to determine the right category. I think that would be HUGELY helpful in understanding how to apply the ALDS.

(Or maybe a demo version where we did the same thing, if you must save the practice for last?)

- 2. On Day 1, I loved getting to do the prompts as a student--so helpful! But then trying to judge our own via the rubric afterwards was very awkward. Of course we all though we did great. More fruitful perhaps to judge each other's, or a sample?

3. I saw a lot of people made cheat sheets where they turned the ALDS into a bulleted grid, which made it easier to compare on certain qualities ("does this have a few details or some? are they well-chosen?") Maybe just give it to us in that form?

You guys all seemed lovely and smart and thoughtful and are doing such good work! But you might need to slow it down a hair for all us non-statisticians.

- I was never sure if I was giving the correct Achievement Levels to the correct papers. If I felt more confident in that I may have had more to say with in the group.

- none

- None.

- none

### **Technical Support in Psychometrics, Assessment Development, and Preparedness for Postsecondary Endeavors**

The Governing Board has a need for technical support to implement some of the technical activities included in the Strategic Vision, such as: planning and designing statistical linking studies; researching how NAEP is used by various audiences and the extent to which various uses are intended and appropriate; developing approaches to updating assessment frameworks while maintaining trends; exploring options to reconfigure the Long-Term Trend assessments; learning best practices from other assessment programs; and exploring the use of NAEP as a measure of preparedness for postsecondary education and careers.

In addition, the [Governing Board's response to the recent evaluation of NAEP achievement levels \(National Academies of Sciences, Engineering, and Medicine; November 2016\)](#) refers to several follow up activities to be pursued over the next few years. Technical support will be needed to assist the Governing Board with implementing some of the activities such as: providing input to inform Board policy revision on setting achievement levels for NAEP; conducting research on the relationship between NAEP achievement levels and concurrent or future performance on measures external to NAEP; conducting research on how the achievement levels are used by various audiences; preparing validity arguments to support the intended uses and interpretations of the achievement levels; and conducting research and producing guidance on inferences best made with achievement levels and those best made with scale scores.

The technical support will include research studies; technical memos; literature reviews and syntheses of best practices; attendance at Governing Board and other meetings; expert consultant services (including the convening of panels); and other ad hoc and quick turnaround requests.

On April 28<sup>th</sup>, a Request for Proposals (RFP) was issued on [www.fbo.gov](http://www.fbo.gov): See [https://www.fbo.gov/index?s=opportunity&mode=form&id=848d447737d77b534c672c5cd1846088&tab=core&\\_cv=1](https://www.fbo.gov/index?s=opportunity&mode=form&id=848d447737d77b534c672c5cd1846088&tab=core&_cv=1)

Proposals were due on June 14<sup>th</sup>, and the proposal review process is currently underway with an anticipated award date of August 2017. The contract period of performance is 12 months for the base year, with annual options for two years.