

National Assessment Governing Board Committee on Standards, Design and Methodology

Friday, November 17, 2017

12:45 – 3:15 pm

AGENDA

12:45 – 1:10 pm	Welcome, Introductions, and Review of Agenda Overview of COSDAM Priorities <i>Andrew Ho, COSDAM Chair</i>	
1:10 – 1:50 pm	Overview of Technical Support Contract (SV #2-10) Discussion of Research Study on Uses of NAEP (SV #3) <i>Thanos Patelis, HumRRO</i>	Attachment A
1:50 – 2:20 pm	Initial Considerations for Combining Assessments (SV #9) <i>Peggy Carr, NCES</i>	Attachment B
2:20 – 2:30 pm	Update on Implementing the Strategic Vision (SV #2-10) Questions on Information Items (See below) <i>Sharyn Rosenberg, Assistant Director for Psychometrics</i>	Attachment C
2:30 – 3:15 pm	CLOSED SESSION : Update on Maintaining Trends with Transition to Digital-Based Assessments (DBA) <i>John Mazzeo, Educational Testing Service</i>	Attachment D
Information Items	Writing Grade 4 Achievement Levels Setting Project Update Update on Implementing the Board's Response to the Evaluation of NAEP Achievement Levels (SV #5) Update on NAEP Validity Studies Panel White Paper (SV #3)	Attachment E Attachment F Attachment G

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

Purpose: The purpose of this session is to provide an update to the Committee on Standards, Design and Methodology (COSDAM) regarding relevant work performed under the Technical Support contract. In the November COSDAM session, project director Dr. Thanos Patelis of the Human Resources Research Organization (HumRRO) will focus primarily on presenting some initial ideas for approaching the research study for COSDAM consideration and discussion.

Project Overview: On August 15, 2017, the National Assessment Governing Board (Governing Board) awarded a contract to HumRRO (as a result of a competitive bidding process) to provide support for technical activities. The Governing Board has a need for such support to implement some of the more technical activities included in the Strategic Vision. 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 activities to be pursued over the next few years. The Technical Support contract 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. This contract will support some of the work performed by COSDAM, the Assessment Development Committee (ADC), and the Ad hoc Committee on Measures of Postsecondary Preparedness. The contract period of performance is 12 months for the base year, with annual options for two years.

Dr. Thanos Patelis is the project director at HumRRO. Key staff members who will be actively contributing to the effort in a number of roles include Dr. Sunny Becker, Dr. Jing Chen, Dr. Monica Gribben, Dr. Arthur Thacker, and Ms. Anne Woods. Subcontractors include EdMetric, LLC and the National Center for the Improvement of Educational Assessment. On the Governing Board staff, Munira Mwalimu is the Contracting Officer and Sharyn Rosenberg is the Contracting Officer's Representative. Lily Clark serves as the technical point of contact for work related to postsecondary preparedness.

November 2017 COSDAM Update:

The following work is currently underway related to COSDAM activities:

Research study on uses of NAEP

The statement of work indicated that the research study topic for the first contract year will be an analysis of how NAEP data in general and achievement levels in particular are used (both appropriately and inappropriately) by various audiences, including but not limited to: policymakers (federal, state, and local); media; researchers; businesses; parents; and the general

public. There are four primary goals to gaining a better understanding of how NAEP results and achievement levels are used: 1) inform a future Board policy statement on intended uses and appropriate interpretations of NAEP; 2) gain understanding of which uses have been most impactful; 3) inform the development of validity arguments for the intended uses and interpretations of NAEP achievement levels, in particular; and 4) inform future reporting and dissemination efforts for increasing appropriate and impactful uses and decreasing inappropriate uses of NAEP results and achievement levels.

During the November COSDAM meeting, project director Thanos Patelis will share some preliminary ideas for approaching the research study.

Literature review on best practices for developing and updating achievement level descriptions

Dr. Karla Egan of EdMetric, LLC will review the literature on the development, use and future of achievement level descriptions (ALDs). In this review, ALDs will be defined, an exploration of the history of ALDs with NAEP will be provided (up to the recent evaluation of NAEP achievement levels), the potential uses of ALDs will be described, and finally the development of valid and aligned ALDs will be explained. The literature review will be available in time for the March 2018 COSDAM meeting, to inform the committee's discussion on revising the Board policy on setting achievement levels for NAEP. Dr. Egan intends to attend the March 2018 COSDAM meeting.

Technical memo on developing a validity argument for the NAEP achievement levels

Dr. Arthur Thacker will review the NAEP achievement levels and all information from existing documents and research reports. Based on this documentation, Dr. Thacker will identify the inferences that can reasonably and appropriately be made from NAEP achievement levels. Additionally, from the gathered information, Dr. Thacker will organize the claims and evidence supporting the validity of the NAEP achievement levels using both a graphical display and in the form of a narrative. The evidence will be organized by claim with a summary of the judgments that the evidence support. Recommendations will be offered for additional research, revisions, or other actions to bolster the evidence supporting the validity of the NAEP achievement levels. The technical memo will be available in time for the March 2018 COSDAM meeting, to inform the committee's discussion on revising the Board policy on setting achievement levels for NAEP. Dr. Thacker intends to attend the March 2018 COSDAM meeting.

Expert panel meeting on NAEP achievement levels setting

A two-day expert panel meeting will take place on January 10-11, 2018 to discuss and provide input to the Governing Board regarding best practices for setting and maintaining achievement levels. This panel meeting is intended to inform the revision of the Board policy on setting achievement levels for NAEP. Eight invited panelists will meet at HumRRO headquarters in Alexandria, Virginia. Several documents will be provided as background reading including: the

current Governing Board policy on setting achievement levels, *the National Academies of Sciences, Engineering, and Medicine 2016 Evaluation of NAEP Achievement Levels*; and the Governing Board's response to the report's recommendations. Input from this expert panel will be documented and included in the March 2018 COSDAM materials to inform the committee's discussion. The list of participating experts is in the process of being finalized and will be shared with COSDAM during the November meeting.

Initial Considerations for Combining or Integrating NAEP Assessments

One of the priorities of the National Assessment Governing Board according to its Strategic Vision is to “Develop policy approaches to revise the NAEP assessment subjects and schedule based on the nation’s evolving needs, the Board priorities, and NAEP funding”. Under this priority, COSDAM plans to “evaluate the technical implications of combining assessments, including the impact on scaling and trends” with stakeholder input. In this session, we will discuss different ways NAEP assessments can be combined and challenges and potential benefits associated with each.

Six issues the Governing Board and NCES must work through in their consideration of this work are:

- Timelines
- Sampling
- Instrumentation and Administration
- Scoring, Scaling and Reporting
- Budgets
- Frameworks and Items

Each of these issues is complex, and requires a good deal of planning, preparation and collaboration. For example, integrating NAEP assessments will require careful review and consideration of existing Frameworks. If a new construct is being formulated by “combining” existing assessments, a elaborated or even new framework would be needed in order to define the content and boundaries of the new construct. Once a combined assessment is formulated, the next step is to design the assessment in a way that it yields reliable and valid scores. Data collection methods, in terms of sampling of students and assignment of blocks and forms to the students, will need to be carefully planned and executed in order to be able to generate reliable and valid scores.

The six issues outlined here will be elaborated and discussed at greater length with COSDAM.

National Assessment Governing Board
Strategic Vision – Implementation Activities Report
November 2017

On November 18, 2016, the National Assessment Governing Board unanimously adopted its Strategic Vision to focus the Board’s work from 2017-2020. This approval marked the beginning of the implementation phase, which is managed by the staff, overseen by the Board, and conducted in partnership with the National Center for Education Statistics (NCES). For each of the four years during the Strategic Vision’s implementation, the staff will provide annual progress reports to the Governing Board at the August Board meetings. This first annual progress report presented in August 2017 was backward-looking, capturing notable work of the past year to implement the Board’s vision.

Beginning with the November 2017 Board meeting, staff are providing the following implementation activities report to identify the things either underway or currently planned to achieve the ten priorities in the Strategic Vision. This report is a living document and will be updated and enhanced based on Board feedback and project decisions; a version of this report will be included in the materials for each quarterly Board meeting for the duration of the Strategic Vision’s implementation. It is included in each standing committee’s materials to provide a snapshot of the entire work plan, to supplement the more detailed committee-specific progress reports created by staff.

National Assessment Governing Board Strategic Vision Implementation Activities Report*

Task Name	Start	Finish	Committee
Strategic Vision	August 4, 2016	March 31, 2025	
SV1 Develop and Sustain Partnerships	November 18, 2016	December 31, 2020	
Work with Partners to Increase Awareness and Use of NAEP	August 4, 2016	December 31, 2020	R&D
Maintain Database of Points of Contact	October 12, 2017	December 31, 2020	R&D
Disseminate Content with/through Partners	October 1, 2016	December 31, 2020	R&D
TUDA Task Force	December 1, 2017	October 16, 2020	Executive Committee
State Policy Task Force	August 8, 2016	August 31, 2020	Executive Committee
SV2 Linking Data	November 18, 2016	December 31, 2020	
Incorporate Ongoing Linking Studies and Consider Additional Work	November 18, 2016	December 31, 2020	COSDAM, NCES
Expand NAEP Linkages to Administrative Data	September 8, 2017	September 8, 2017	NCES
Board Considers What Federal Data Presented with NAEP	September 8, 2017	December 5, 2018	R&D
Learn from Reporting of International Assessments	November 17, 2017	November 18, 2017	R&D
SV3 Expand NAEP Resources	November 18, 2016	December 31, 2020	
Create Tools for New Audiences (also SV4)	April 3, 2018	April 3, 2019	R&D
Develop 'Menu of Engagement'	January 1, 2018	December 31, 2020	R&D
Create 'Brief Case' Studies on NAEP Use (also SV4)	January 1, 2018	December 31, 2020	R&D
Build Teacher Prep Toolkit	August 2, 2018	August 9, 2019	R&D
Share Effective Uses of NAEP	March 1, 2018	October 31, 2018	R&D
Research Uses of NAEP by Various Audiences	March 3, 2017	November 16, 2018	COSDAM, R&D
Improve Understanding of NAEP Achievement Levels	October 12, 2017	December 31, 2020	R&D, COSDAM
Develop Statement of Intended and Appropriate Uses of NAEP	November 16, 2018	May 18, 2019	COSDAM
Host Stakeholder Panels at Board Meetings	August 8, 2016	December 31, 2020	
Disseminate Information on NAEP Technical Procedures to Share Expertise	March 1, 2018	December 31, 2020	COSDAM, NCES
Identify NAEP Resources & Information for Practitioners (also SV6)	May 18, 2017	March 18, 2021	ADC, R&D
SV4 Dissemination and Use of NAEP	November 18, 2016	December 31, 2020	
Post-release Stakeholder Events to Extend Life of Results	April 2, 2018	December 31, 2020	R&D
Update Governing Board Website	October 3, 2016	July 24, 2017	R&D
Expand Capability for More Wide-ranging Communications Approaches and Products	October 12, 2017	December 31, 2020	R&D
Identify Advanced and More User-friendly Approaches to Presenting NAEP Results	October 12, 2017	December 31, 2020	R&D
SV5 Update Frameworks	November 18, 2016	December 31, 2020	
Update Framework Development Policy	June 5, 2017	March 3, 2018	ADC

National Assessment Governing Board Strategic Vision Implementation Activities Report*

Task Name	Start	Finish	Committee
Update Item Development Policy	March 1, 2018	November 14, 2018	ADC
Explore New Approaches to Framework Update Processes (also SV8)	November 17, 2017	March 1, 2019	ADC
Review & Update Reading Framework (concurrent w/ Math)	October 9, 2017	March 31, 2025	ADC
Review & Update Mathematics Framework (concurrent w/ Reading)	June 30, 2017	March 31, 2025	ADC
Review & Update Civics, Geography, and U.S. History Frameworks	May 1, 2018	May 15, 2020	ADC
Review & Update Economics Framework (Depends on Board Decision on Social Studies)	March 6, 2020	August 6, 2021	ADC
Review & Update Science and TEL Frameworks	September 1, 2020	November 18, 2022	ADC
Review & Update Writing Framework	March 7, 2022	August 4, 2023	ADC
Update Board Policy on Achievement Levels (Including New Approaches to ALDs)	January 2, 2017	August 4, 2018	COSDAM
SV6 Contextual Variables	November 18, 2016	December 31, 2020	
R&D Review	August 4, 2017	December 30, 2019	R&D
SV7 Long-Term Trend	August 8, 2016	May 18, 2018	
Ed Haertel Overview Paper	August 8, 2016	December 9, 2016	COSDAM
Reaction Papers (4)	December 12, 2016	February 17, 2017	COSDAM
Washington DC Symposium	February 20, 2017	February 20, 2017	Full Board
AERA Symposium	February 20, 2017	February 20, 2017	Full Board
Governing Board Discussions	March 3, 2017	May 18, 2018	Full Board
Governing Board Action	May 18, 2018	May 18, 2018	Full Board
SV8 Other Countries	November 17, 2017	November 17, 2017	
International Assessment Expert Panel	November 17, 2017	November 17, 2017	Full Board
SV9 Assessment Schedule	May 19, 2017	September 24, 2018	
Develop Policy Priorities	May 19, 2017	March 2, 2018	Executive Committee
Review Technical Implications of Combining Assessments	November 17, 2017	May 18, 2018	COSDAM
Revise NAEP Assessment Schedule	March 5, 2018	September 24, 2018	Executive Committee
SV10 Postsecondary Preparedness	August 6, 2016	August 31, 2020	
Ad Hoc Committee Develops Recommendations	August 3, 2017	November 17, 2018	Ad Hoc Committee
Implement Approved Recommendations of Ad Hoc Committee	November 19, 2018	August 31, 2020	Full Board
Continue Research to Gather Validity Evidence on Academic Preparedness for College	August 6, 2016	August 31, 2020	COSDAM

Strategic Vision Activities Led by COSDAM

The previous document lists high level activities intended to support each of the ten Strategic Vision priorities. For each activity led by COSDAM, additional information is provided below to describe the current status and recent work, planned next steps, and the ultimate desired outcomes. 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 standing committee (e.g., framework update processes) or ad hoc committee (i.e., exploring new approaches to postsecondary preparedness) also have not been included below.

During this brief update, the Governing Board's Assistant Director for Psychometrics, Sharyn Rosenberg, will highlight a few recent developments and will answer any questions that COSDAM members have about ongoing or planned activities.

Strategic Vision Activity	Current Status and Recent Work	Planned Next Steps	Desired Outcome
SV #2: Increase opportunities to connect NAEP to administrative data and state, national, and international student assessments <i>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>	COSDAM discussions at May and August 2017 board meetings to examine how existing findings may be used to add meaning to scale scores and achievement levels, and what additional studies to take on 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	Informational update on current studies will be provided in the March 2018 COSDAM materials Complete ongoing studies Decide what new studies to take on Decide how to use and report existing and future results Complete additional studies	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)

Strategic Vision Activity	Current Status and Recent Work	Planned Next Steps	Desired Outcome
<p>SV #3: Expand the availability, utility, and use of NAEP resources, in part by creating new resources to inform education policy and practice</p> <p><i>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&D and ADC)</i></p> <p><i>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><i>Disseminate information on technical best practices and NAEP methodologies, such as training item writers and setting achievement levels</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>Technical Support contract specifies that the research study topic for year 1 will focus on how NAEP results are used by various stakeholders. Initial ideas for approaching this research study will be shared with COSDAM during the November 2017 meeting</p> <p>This idea was generated during the August 2017 COSDAM discussion of the Strategic Vision activities</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>Work with NCES and R&D to refine list of technical topics for dissemination efforts</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 the policy statement)</p> <p>Stakeholders benefit from NAEP technical expertise</p>

Strategic Vision Activity	Current Status and Recent Work	Planned Next Steps	Desired Outcome
<p>SV# 5: Develop new approaches to update NAEP subject area frameworks to support the Board's responsibility to measure evolving expectations for students, while maintaining rigorous methods that support reporting student achievement trends</p> <p><i>Consider new approaches to creating and updating the achievement level descriptors and update the Board policy on achievement levels</i></p>	<p>Panel of standard setting experts to discuss technical issues and recommendations for achievement levels policy (January 2018)</p> <p>Literature review/synthesis of best practices for creating and updating achievement level descriptors (ALDs) (February 2018)</p> <p>Technical Memo on developing a validity argument for the NAEP achievement levels (February 2018)</p>	<p>COSDAM discussion of draft policy statement and supporting materials (March 2018)</p> <p>Seek external feedback and public comment (April 2018)</p> <p>Revised policy statement for full Board discussion (May 2018)</p> <p>Board action on revised policy statement (August 2018)</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>SV# 7: Research policy and technical implications related to the future of NAEP Long-Term Trend assessments in reading and mathematics</p> <p><i>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>Full Board and Executive Committee discussions (March, May, and August 2017) and webinar on secure LTT items and p-values from 2012 administration (October 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 and/or whether changes to the design and administration of the LTT assessment are needed (led by Executive Committee and NCES)</p>

Strategic Vision Activity	Current Status and Recent Work	Planned Next Steps	Desired Outcome
<p>SV# 9: Develop policy approaches to revise the NAEP assessment subjects and schedule based on the nation's evolving needs, the Board's priorities, and NAEP funding</p> <p><i>Pending outcomes of stakeholder input (ADC activity), evaluate the technical implications of combining assessments, including the impact on scaling and trends</i></p>	<p>Presentation and discussion on initial considerations for combining assessments (November 2017)</p>	<p>TBD</p>	<p>TBD</p>
<p>SV# 10: Develop new approaches to measure the complex skills required for transition to postsecondary education and career</p> <p><i>Continue research to gather validity evidence for using 12th grade NAEP reading and math results to estimate the percentage of grade 12 students academically prepared for college</i></p>	<p>Several studies are ongoing (see activities under SV# 2)</p> <p>Per COSDAM discussion at August 2017 meeting, additional studies are on hold until at least November 2018 pending Board decision on how to move forward with findings from Ad hoc Committee on Measures of Postsecondary Preparedness</p>	<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>	<p>Statements about using NAEP as an indicator of academic preparedness for college continue to be defensible and to have appropriate validity evidence</p>

Transition to Digital-Based Assessments: 2017 Mode Evaluation Analyses

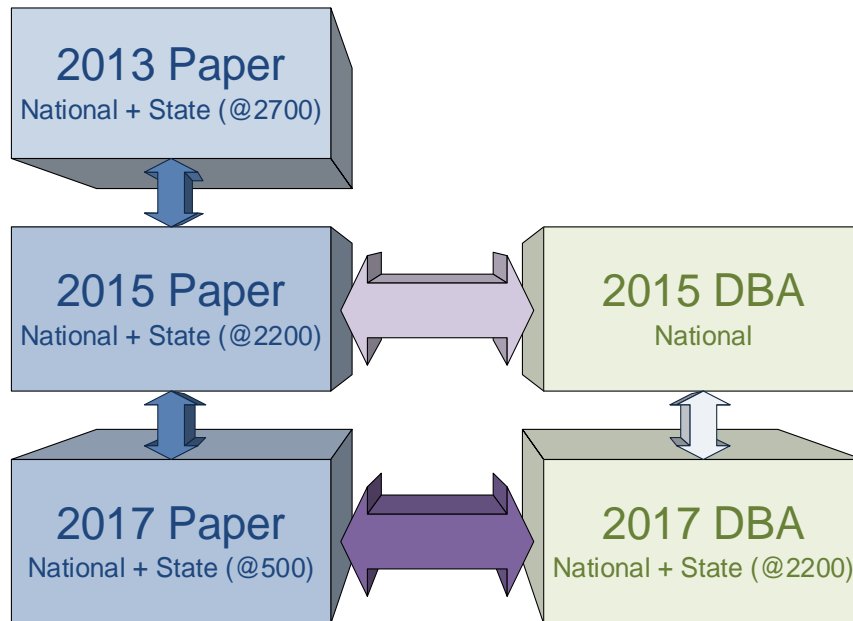
NAEP has always had two complementary, and in some ways competing, goals. NAEP is a measure of academic progress, and thus tracking improvements or declines in student performance is central to the program. Measuring trends normally means changing the assessments as little as possible from one year to the next so that any changes observed can be unambiguously interpreted as due to changes in the tested population. However, NAEP is also the only nationally representative measure of what students know and can do. In this role, it is essential that NAEP measure what students are actually being taught or the knowledge and skills they are expected to have. From this viewpoint, measuring trends with respect to a fixed construct may not be a helpful or policy-relevant goal for the program. Such a viewpoint argues against the use of assessments or administration modes that are dated and have no current relevance, even if those are optimal to the unambiguous interpretation of change.

In 2017, the NAEP program continued the careful, multistep process of transitioning the core assessment in reading and mathematics from paper to digital-based assessments. While the program has been studying and reporting student performance on writing in a digitally based environment since 2011 and in technology and engineering literacy in 2014, the 2017 reading and mathematics assessments will be the first operational assessments in a digital format of these two core subjects. Bridge studies, starting with the 2015 field assessment, have been designed to study the transition from paper-based assessment (PBA) to digital-based assessment (DBA) carefully and, subsequently, create the conditions under which trend reporting can be supported. The initial design was relatively straightforward, adding a tablet-based national component to the operational assessment that could be compared with the paper-based operational assessment in that year. In the 2017 operational assessment, the bridge design from the 2015 field study was expanded to include state- and urban district-level results in both the paper and tablet based assessment modes.

The 2017 reading and mathematics assessments were administered to representative samples of students from the nation overall and from public schools in the states and participating districts. For each subject area and grade level assessed, a sample of approximately 500 students in each jurisdiction were assigned to take the paper-based assessment in the same conditions used in 2015 (and prior years). In addition, each jurisdiction had approximately 2,200 students randomly assigned to take the assessment in its digital format.

During this closed session presentation, committee members will see the results of the 2017 within-year mode evaluation analyses, including observed item statistics and the equated group scale score comparisons between PBA and DBA. The scale score results estimated for the purpose of the mode evaluation will not be identical to those that will be in the final Report Card; however, these results will provide valuable information to further inform NAEP's transition to digital-based assessments.

As discussed at the March 2016 COSDAM meeting, the graphic below provides an overview of the various samples and conditions involved in the linking design, where ‘State’ represents both state and urban districts.



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



Purpose: The purpose of this document 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.

Legend:

Light shading: Completed

No shading: To be completed after 11/16/17

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 2017.¹

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 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):

¹ 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 Cizek²

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

Senior Advisor for Measurement Services, Measured Progress

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)

November 2017 Update:

Update on Preparations for the Pilot Study

Pearson is currently in the process of finalizing all materials, tools, and logistics necessary to conduct the pilot study. All material and tool revisions were based on lessons learned from the June field trial in addition to feedback from TACSS. The pilot study will be conducted from November 6-9, 2017 in Atlanta, GA. Twenty-two panelists from around the country were recruited and have committed to participating.

September TACSS Meeting

On September 14th and 15th, the TACSS met to review and discuss materials and procedural preparations for the pilot study. The majority of the meeting focused on detailed review of draft materials and facilitation notes. Editorial comments and suggestions were collected for all presentations, process evaluations, the ALS overview, and the facilitation guide. Discussion of the exemplar selection activity scheduled on the final day of the study highlighted the fact that panelists would have fewer than 16 bodies of work to evaluate. These bodies of work will be familiar to all panelists and comprised of responses to two prompts chosen for public release. Concern was expressed around the potential that panelists might have few bodies of work to evaluate within a given performance level, depending on where the cut score recommendations fall (in particular within the Advanced performance level). As such, TACSS

² Greg Cizek was appointed to the Governing Board by Secretary Betsy DeVos to serve as one of the three Testing and Measurement experts from October 1, 2017 to September 30, 2021. On October 7, 2017, Greg informed Pearson that he was resigning from the TACSS. Given the project timeline and the small number of remaining TACSS meetings, Pearson will not seek a replacement TACSS member.

recommended that a second stage be added to this activity to allow for additional bodies of work to be considered.

The second stage will involve the selection of additional bodies of work that panelists will not have seen yet. The intention will be to select these such that each performance level ends up with roughly the same overall number of bodies of work to evaluate. If the two stage process does not end up working well, it will be removed for the operational ALS.

Given the proximity of the November 2017 COSDAM session to the end of the pilot study, a separate webinar is planned for early to mid-December to provide a briefing on outcomes. During that webinar, Writing ALS Project Director Tim O'Neil will present results and lessons learned from the pilot study and provide an update on preparations for the operational study.

Update on Implementing the Board's Response to the Evaluation of NAEP Achievement Levels

Overview

On November 17, 2016, the National Academies of Sciences, Engineering, and Medicine released the final report of their evaluation, *Evaluation of the achievement levels for mathematics and reading on the National Assessment of Educational Progress*. A free PDF of the full report can be downloaded at: <https://www.nap.edu/catalog/23409/evaluation-of-the-achievement-levels-for-mathematics-and-reading-on-the-national-assessment-of-educational-progress>. The Governing Board received a briefing from staff at the National Academies of Sciences, Engineering, and Medicine and members of the interdisciplinary review committee during the quarterly Board meeting on November 19, 2016.

As stated in the NAEP legislation, the Commissioner of NCES is to use the findings from the evaluation to decide whether the achievement levels should continue to be used on a “trial basis” or whether that designation can be removed. In addition, the final report included conclusions and recommendations that have implications for future Governing Board achievement levels-setting activities. Public Law 107-279 specifies that the Governing Board must prepare a formal response to the evaluation:

Not later than 90 days after an evaluation of the student achievement levels under section 303(e), the Assessment Board shall make a report to the Secretary, the Committee on Education and the Workforce of the House of Representatives, and the Committee on Health, Education, Labor, and Pensions of the Senate describing the steps the Assessment Board is taking to respond to each of the recommendations contained in such evaluation.

Due to the timing of the evaluation report release, the 90 day window concluded prior to the March 2017 Governing Board meeting. Therefore, on November 19, 2016, the Board granted a joint delegation of authority to COSDAM and the Executive Committee for formal approval of the report to the Secretary, the Committee on Education and the Workforce of the House of Representatives, and the Committee on Health, Education, Labor, and Pensions of the Senate describing the steps the Governing Board is taking to respond to each of the recommendations contained in the evaluation.

COSDAM met via teleconference on December 9, 2016 to discuss an initial draft response to the evaluation. On December 19, 2016, the Executive Committee and COSDAM met to discuss and take action on a revised response. The final response (attached) was approved by a vote of 9-0 with one abstention. The response was sent to Secretary John King, the Committee on Education and the Workforce of the House of Representatives, and the Committee on Health, Education, Labor, and Pensions of the Senate on December 20, 2016.

Ongoing discussions with COSDAM and the full Board began during the March 2017 quarterly meeting to implement the recommendations from the evaluation and have continued at each quarterly meeting of COSDAM.

Status of Implementing the Board's Response

Revision of Board Policy

The first step in implementing the Governing Board response is to update the current policy on [Developing Student Performance Levels for NAEP](#). The revised policy is intended to address several of the issues raised in the evaluation, most notably the need for specifying a process and timeline for conducting regular recurring reviews of the achievement level descriptions and the need to be more explicit about conditions that require a new standard setting. The policy revision will also incorporate other recent developments in standard setting.

There are several ongoing activities to support the policy revision that are being carried out through the Board's Technical Support contract, including a literature review on best practices for developing and updating achievement level descriptions, a panel meeting of standard setting experts, and a technical memo on developing a validity argument for the NAEP achievement levels (see Attachment A). The March 2018 COSDAM meeting will focus on reviewing and discussing these materials and a first draft of a proposed revised policy.

Regular Recurring Review of the Achievement Level Descriptors

Once the Board policy has been revised to incorporate guidance for specifying a process and timeline for conducting regular recurring review of the achievement level descriptions, a statement of work will be developed to procure services for performing this work. Additional details will be discussed with COSDAM during spring/summer 2018.

Relationships Between NAEP Achievement Levels and External Measures

The Strategic Vision includes an explicit goal to increase opportunities for connecting NAEP to other national and international assessments and data, and COSDAM has had several discussions related to this topic over the past year. In particular, the May 2017 COSDAM meeting included a synthesis of findings from existing linking studies as a means of better understanding how NAEP achievement levels and scale scores relate to other important indicators of student performance. Throughout 2018, COSDAM will continue to receive updates and discuss ongoing and planned linking studies to support the goal of making NAEP achievement levels and scale scores more meaningful by connecting them to measures external to NAEP.

Interpretations and Uses of NAEP Achievement Levels and Guidance for Inferences Made with Achievement Levels versus Scale Scores

Much of the Strategic Vision focuses on improving the use and dissemination of NAEP results, including the achievement levels. Under the Technical Support contract, the research study on uses of NAEP (to be completed by August 2018) will synthesize how various audiences have used and interpreted NAEP results, including achievement levels. In addition, a technical memo on developing a validity argument for the NAEP achievement levels (to be completed by February 2018) will help provide a framework for evaluating the validity of common uses of the achievement levels. Governing Board staff will work with NCES to provide additional recommendations for addressing the Board's response to these items, which will be discussed with COSDAM throughout 2018.

National Assessment Governing Board’s Response to the National Academies of Sciences, Engineering, and Medicine 2016 Evaluation of NAEP Achievement Levels

Legislative Authority

Pursuant to the National Assessment of Educational Progress (NAEP) legislation (Public Law 107-279), the National Assessment Governing Board (hereafter the Governing Board) is pleased to have this opportunity to apprise the Secretary of Education and the Congress of the Governing Board response to the recommendations of the National Academies of Sciences, Engineering, and Medicine evaluation of the NAEP achievement levels for mathematics and reading (Edley & Koenig, 2016).

The cited legislation charges the Governing Board with the authority and responsibility to “develop appropriate student achievement levels for each grade or age in each subject area to be tested.” The legislation also states that “such levels shall be determined by... a national consensus approach; used on a trial basis until the Commissioner for Education Statistics determines, as a result of an evaluation under subsection (f), that such levels are reasonable, valid, and informative to the public; ... [and] shall be updated as appropriate by the National Assessment Governing Board in consultation with the Commissioner for Education Statistics” (Public Law 107-279).

Background

NAEP is the largest nationally representative and continuing assessment of what our nation’s elementary and secondary students know and can do. Since 1969, NAEP has been the country’s foremost resource for measuring student progress and identifying differences in student achievement across student subgroups. In a time of changing state standards and assessments, NAEP serves as a trusted resource for parents, teachers, principals, policymakers, and researchers to compare student achievement across states and select large urban districts. NAEP results allow the nation to understand where more work must be done to improve learning among all students.

For 25 years, the NAEP achievement levels (*Basic*, *Proficient*, and *Advanced*) have been a signature feature of NAEP results. While scale scores provide information about student achievement over time and across student groups, achievement levels reflect the extent to which student performance is “good enough,” in each subject and grade, relative to aspirational goals.

Since the Governing Board began setting standards in the early 1990s, achievement levels have become a standard part of score reporting for many other assessment programs in the US and abroad.

Governing Board Response

Overview

The Governing Board appreciates the thorough, deliberative process undertaken over the past two years by the National Academies of Science, Engineering, and Medicine and the expert members of the Committee on the Evaluation of NAEP Achievement Levels for Mathematics and Reading. The Governing Board is pleased that the report concludes that the achievement levels are a meaningful and important part of NAEP reporting. The report states that, “during their 24 years [the achievement levels] have acquired meaning for NAEP’s various audiences and stakeholders; they serve as stable benchmarks for monitoring achievement trends, and they are widely used to inform public discourse and policy decisions. Users regard them as a regular, permanent feature of the NAEP reports” (Edley & Koenig, 2016; page Sum-8). The Governing Board has reviewed the seven recommendations presented in the report and finds them reasonable and thoughtful. The report will inform the Board’s future efforts to set achievement levels and communicate the meaning of NAEP *Basic*, *Proficient*, and *Advanced*. The recommendations intersect with two Governing Board documents, the Strategic Vision and the achievement levels policy, described here.

On November 18, 2016, the Governing Board adopted a Strategic Vision (<https://www.nagb.org/content/nagb/assets/documents/newsroom/press-releases/2016/nagb-strategic-vision.pdf>) to guide the work of the Board through 2020, with an emphasis on innovating to enhance NAEP’s form and content and expanding NAEP’s dissemination and use. The Strategic Vision answers the question, “How can NAEP provide information about how our students are doing in the most innovative, informative, and impactful ways?” The Governing Board is pleased that several of the report recommendations are consistent with the Board’s own vision. The Governing Board is committed to measuring the progress of our nation’s students toward their acquisition of academic knowledge, skills, and abilities relevant to this contemporary era.

The Governing Board’s approach to setting achievement levels is articulated in a policy statement, “Developing Student Performance Levels for the National Assessment of Educational Progress” (<https://www.nagb.org/content/nagb/assets/documents/policies/developing-student-performance.pdf>). The policy was first adopted in 1990 and was subsequently revised in 1995, with minor wording changes made in 2007. The report motivates the revision of this policy, to add clarity and intentionality to the setting and communication of NAEP achievement levels.

The seven recommendations and the Governing Board response comprise a significant research and outreach trajectory that the Governing Board can pursue over several years in conjunction with key partners. The Governing Board will implement these responses within resource constraints and in conjunction with the priorities of the Strategic Vision.

Evaluating the Alignment of NAEP Achievement Level Descriptors

Recommendation #1: Alignment among the frameworks, the item pools, the achievement-level descriptors, and the cut scores is fundamental to the validity of inferences about student achievement. In 2009, alignment was evaluated for all grades in reading and for grade 12 in mathematics, and changes were made to the achievement-level descriptors, as needed. Similar research is needed to evaluate alignment for the grade 4 and grade 8 mathematics assessments and to revise them as needed to ensure that they represent the knowledge and skills of students at each achievement level. Moreover, additional work to verify alignment for grade 4 reading and grade 12 mathematics is needed.

The report's primary recommendation is to evaluate the alignment, and revise if needed, the achievement level descriptors for NAEP mathematics and reading assessments in grades 4, 8, and 12. The Governing Board intends to issue a procurement for conducting studies to achieve this goal. The Governing Board has periodically conducted studies to evaluate whether the achievement level descriptors in a given subject should be revised, based on their alignment with the NAEP framework, item pool, and cut scores. The Governing Board agrees that this is a good time to ensure that current NAEP mathematics and reading achievement level descriptors align with the knowledge and skills of students in each achievement level category. In conjunction with the response to Recommendation #3, the updated Board policy on NAEP achievement levels will address the larger issue of specifying a process and timeline for conducting regular recurring reviews of the achievement level descriptions in all subjects and grades.

The Governing Board agrees strongly with the recommendation that, while evaluating alignment of achievement level descriptors is timely, it is not necessary to consider changing the cut scores or beginning a new trend line at this time. The NAEP assessments are transitioning from paper-based to digital assessments in 2017, and current efforts are focused on ensuring comparability between 2015 and 2017 scores. The Governing Board articulated this in the 2015 Resolution on Maintaining NAEP Trends with the Transition to Digital-Based Assessments (<https://www.nagb.org/content/nagb/assets/documents/policies/resolution-on-trend-and-dba.pdf>).

Recommendation #2: Once satisfactory alignment among the frameworks, the item pools, the achievement-level descriptors, and the cut scores in NAEP mathematics and reading has been demonstrated, their designation as trial should be discontinued. This work should be completed and the results evaluated as stipulated by law: (20 U.S. Code 9622: National Assessment of Educational Progress: <https://www.law.cornell.edu/uscode/text/20/9622> [September 2016]).

Ultimately, the Commissioner of Education Statistics is responsible for determining whether the “trial” designation is removed. The Governing Board is committed to providing the Commissioner with the information needed to make this determination in an expedient manner.

Regular Recurring Reviews of the Achievement Level Descriptors

Recommendation #3: To maintain the validity and usefulness of achievement levels, there should be regular recurring reviews of the achievement-level descriptors, with updates as needed, to ensure they reflect both the frameworks and the incorporation of those frameworks in NAEP assessments.

The Board’s current policy on NAEP achievement levels contains several principles and guidelines for *setting* achievement levels but does not address issues related to the continued use or reporting of achievement levels many years after they were established. The revised policy will seek to address this gap by including a statement of periodicity for conducting regular recurring reviews of the achievement level descriptors, with updates as needed, as called for in this recommendation. The Governing Board agrees that it is important to articulate a process and timeline for conducting regular reviews of the achievement level descriptors rather than performing such reviews on an ad hoc basis.

Relationships Between NAEP Achievement Levels and External Measures

Recommendation #4: Research is needed on the relationships between the NAEP achievement levels and concurrent or future performance on measures external to NAEP. Like the research that led to setting scale scores that represent academic preparedness for college, new research should focus on other measures of future performance, such as being on track for a college-ready high school diploma for 8th-grade students and readiness for middle school for 4th-grade students.

In addition to the extensive work that the Governing Board has conducted at grade 12 to relate NAEP mathematics and reading results to academic preparedness for college, the Governing Board has begun research at grade 8 with statistical linking studies of NAEP mathematics and reading and the ACT Explore assessments in those subjects. This work was published while the evaluation was in process and was not included in the Committee’s deliberations. Additional studies in NAEP mathematics and reading at grades 4 and 8 are beginning under contract to the National Center for Education Statistics (NCES). The Governing Board’s Strategic Vision includes an explicit goal to increase opportunities for connecting NAEP to other national and international assessments and data. Just as the Board’s previous research related grade 12 NAEP results in mathematics and reading to students’ academic preparedness for college, the Governing Board anticipates that additional linkages with external measures will help connect

the NAEP achievement levels and scale scores to other meaningful real-world indicators of current and future performance.

Interpretations and Uses of NAEP Achievement Levels

Recommendation #5: Research is needed to articulate the intended interpretations and uses of the achievement levels and collect validity evidence to support these interpretations and uses. In addition, research to identify the actual interpretations and uses commonly made by NAEP's various audiences and evaluate the validity of each of them. This information should be communicated to users with clear guidance on substantiated and unsubstantiated interpretations.

The Governing Board's Strategic Vision emphasizes improving the use and dissemination of NAEP results, and the Board's work in this area will include achievement levels. The Governing Board recognizes that clarity and meaning of NAEP achievement levels (and scale scores) are of utmost importance. The Governing Board will issue a procurement to conduct research to better understand how various audiences have used and interpreted NAEP results (including achievement levels). The Governing Board will work collaboratively with NCES to provide further guidance and outreach about appropriate and inappropriate uses of NAEP achievement levels.

Guidance for Inferences Made with Achievement Levels versus Scale Scores

Recommendation #6: Guidance is needed to help users determine inferences that are best made with achievement levels and those best made with scale score statistics. Such guidance should be incorporated in every report that includes achievement levels.

The Governing Board understands that improper uses of achievement level statistics are widespread in the public domain and extend far beyond the use of NAEP data. Reports by the Governing Board and NCES have modeled appropriate use of NAEP data and will continue to do so. This recommendation is also consistent with the goal of the Strategic Vision to improve the dissemination and use of NAEP results. The Governing Board will continue to work with NCES and follow current research to provide guidance about inferences that are best made with achievement levels and those best made with scale score statistics.

Regular Cycle for Considering Desirability of Conducting a New Standard Setting

Recommendation #7: NAEP should implement a regular cycle for considering the desirability of conducting a new standard setting. Factors to consider include, but are not limited to: substantive changes in the constructs, item types, or frameworks; innovations in the modality for administering assessments; advances in standard setting methodologies; and changes in the

policy environment for using NAEP results. These factors should be weighed against the downsides of interrupting the trend data and information.

When the Board's achievement levels policy was first created and revised in the 1990s, the Board was setting standards in each subject and grade for the first time and had not yet considered the need or timeline for re-setting standards. To address this recommendation, the Governing Board will update the policy to be more explicit about conditions that require a new standard setting.

Board's Commitment

The Governing Board remains committed to its congressional mandate to set “appropriate student achievement levels” for the National Assessment of Educational Progress. The Board appreciates the report's affirmation that NAEP achievement levels have been set thoughtfully and carefully, consistent with professional guidelines for standard setting, and based on extensive technical advice from respected psychometricians and measurement specialists. The Board also takes seriously the charge to develop the current achievement levels through a national consensus approach, involving large numbers of knowledgeable teachers, curriculum specialists, business leaders, and members of the general public throughout the process. This is only fitting given the Governing Board's own congressionally mandated membership that explicitly includes representatives from these stakeholder groups.

The Governing Board remains committed to improving the process of setting and communicating achievement levels. The Governing Board is grateful for the report recommendations that will advance these aims.

Reference

Edley, C. & Koenig, J. A. (Ed.). (2016). *Evaluation of the Achievement Levels for Mathematics and Reading on the National Assessment of Educational Progress*. Washington, DC: The National Academies Press.

White Paper on Uses and Misuses of NAEP Data

The NAEP Validity Studies Panel (NVS) is developing a paper to inform a concise and modest statement of the intended uses and interpretations of NAEP scores. This request was based on the observation that the first standard from the AERA/APA/NCME standards is, "...the test developer should set forth clearly how test scores are intended to be interpreted and consequently used." The paper will discuss a variety of intended interpretations and uses that are explicit and implicit across NAEP legislation and products. It will also discuss interpretations and uses that may not be intended but are nonetheless widespread.

After meeting with the COSDAM panel in March of this year, additional sources were identified for considerations which are currently being incorporated into the white paper. Finally, the panel underscored the vastness this paper could potentially take on. In that respect, a conversation at the next NVS panel meeting will include discussion on how to limit the scope while still maintaining the original goal of this project.

A draft of the white paper should be available to NCES in the Fall of 2018.