Committee on Standards, Design and Methodology

May 3, 2021

12:00 – 2:00 pm ET (Virtual)

AGENDA

12:00 – 12:05 pm	**Welcome and Overview of Agenda** *Gregory Cizek, Chair*	
12:05 – 12:20 pm	**Update: Review and Revision of Mathematics and Reading Achievement Level Descriptions** *Eric Moyer, Pearson* *Sharyn Rosenberg, Governing Board staff*	Attachment A
12:20 – 12:35 pm	**Update: Framework Development Processes** *Gregory Cizek*	Attachment B
12:35 – 1:30 pm	**Discussion of NAEP Reading Assessment** *Gregory Cizek*	Attachment C
1:30 – 2:00 pm	**Discussion of 2022 NAEP Long-Term Trend** *Enis Dogan, National Center for Education Statistics* *Gregory Cizek*	Attachment D

Studies to Review and Revise NAEP Achievement Level Descriptions (ALDs) for Mathematics, Reading, and Other Subjects

Background

On September 24, 2020, the National Assessment Governing Board (Governing Board) awarded contract# 91995920C0004 to Pearson (as a result of a competitive bidding process) for conducting studies to review and revise NAEP achievement level descriptions (ALDs) in mathematics and reading using the 2019 NAEP assessments at grades 4, 8, and 12[1]. This work is intended to address the first recommendation of the evaluation of NAEP achievement levels that was conducted by the National Academies of Sciences, Engineering, and Medicine:

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 Board committed to conducting studies to review and revise the NAEP ALDs in its initial response to the evaluation that was formally adopted and sent to the Secretary of Education and Congress in December 2016. The Board's Achievement Levels Work Plan, adopted in March 2020, further describes the intention for this work: "Addressing Recommendation #1 should focus on the current reporting ALDs for mathematics and reading at grades 4, 8, and 12. The methodology will be similar to what was done to evaluate the alignment and revise the 2009 NAEP Reading ALDs for grades 4, 8, and 12 (Donohue, Pitoniak, & Beaulieu, 2010) and the 2009 NAEP Mathematics ALDs for grade 12 (Pitoniak, Dion, & Garber, 2010). This process will generate new reporting ALDs that comply with the revised Board policy statement" (p. 3).

According to Principle 1a of the Board policy on Developing Student Achievement Levels for NAEP, "Content achievement level descriptions translate the policy definitions into specific

[1] The base period of this contract includes the review and revision of ALDs in mathematics and reading at grades 4, 8, and 12; in addition, an option may be exercised for a second phase of the contract focusing on review and revision of ALDs in U.S. history, civics, science, technology and engineering literacy (TEL) at grade 8 based on data from the most recent administrations of those assessments in 2018 and 2019.

expectations about student knowledge and skills in a particular content area, at each achievement level, for each subject and grade. Content ALDs provide descriptions of specific expected knowledge, skills, or abilities of students performing at each achievement level. They reflect the range of performance that items and tasks should measure. When setting achievement levels, the content ALDs provide consistency and specificity for panelist interpretations of policy definitions for a given assessment. During reporting, content ALDs communicate the specific knowledge and skills represented by *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* for a given assessment" (p. 5).

Principles 3g and 4a of the Board policy apply specifically to this project of reviewing and revising the current ALDs and creating reporting ALDs (based on empirical data) that indicate what students at each achievement level *do* know and *can* do rather than what they *should* know and *should* be able to do[2]. Additional details for carrying out the work described by principles 3g and 4a are included in the Achievement Levels Procedures Manual.

The basis for the evaluation of NAEP achievement levels (and subsequently for this project) is the *existing* NAEP frameworks and item pools, not the new NAEP Mathematics Framework currently scheduled for implementation in 2025 or the NAEP Reading Framework that is currently under development and consideration by the Board. In accordance with principle 4b of the Board policy, the achievement levels and/or ALDs will need to be reviewed again once the new frameworks are implemented. Such work is beyond the scope of this project.

Project Overview

Dr. Eric Moyer is the project director at Pearson and Dr. Jennifer Galindo is the assistant project director at Pearson. Pearson will conduct a pilot study and an operational meeting using scale anchoring studies where panels of content experts judge the alignment of the current mathematics and reading ALDs and produce a set of recommended reporting ALDs for the Governing Board to consider in reporting the results from the next regular administration of the NAEP reading and mathematics assessments at grades 4, 8, and 12. The Governing Board is expected to take action on the reporting ALDs for mathematics and reading at grades 4, 8, and 12 in advance of the next release of these results.

Based on careful review of the history of ALD development, review, and revisions for NAEP mathematics and reading, a model-based anchored approach for reviewing the alignment of the ALDs for NAEP mathematics and reading will be used. The methodology for this alignment review study is based on that of previous studies, including the ALD development and review meeting held in 2009. The methodology was specified by the Board's Achievement Levels Work Plan and was selected to reduce the potential for possible inconsistencies from the use of different methods. The process of the model-based anchored approach will result in organizing

[2] According to the Board policy, ALDs will continue to describe what students *should* know and *should* be able to do for the purposes of item development and standard setting; only the reporting ALDs will be written in terms of what students *do* know and *can* do.

specific NAEP items by achievement level, which will serve as a key referent for panelists in reviewing and revising the current ALDs.

The model-based anchored approach includes three stages. The first stage will involve conducting statistical analyses to determine the items from the subject and grade that are anchored to a level corresponding to the score range within cut scores set to represent the achievement level descriptors (ALDs). The second stage relies on panels of content experts for each individual assessment. The panelists individually review the items that are anchored to each performance level and create summary descriptions of what students in each level are expected to know and be able to demonstrate based on the knowledge and skills measured by the items. In the final stage, the panelists compare the current ALDs for the respective assessment with their summary descriptions. The panelists note the similarities and differences, to make a recommendation regarding whether the current ALDs accurately describe what students in each level are expected to know and be able to demonstrate or if revisions to the current ALDs are needed to improve alignment. The final alignment judgment will be used to report whether the panels determined that there exists alignment between the current ALDs and student expectations. The final panel summary descriptions will be used to revise the current ALDs to create reporting ALDs that indicate what students at each achievement level do know and can do.

There is a technical advisory committee (TAC) consisting of the following experts in ALDs:

Dr. Karla Egan (Principal, EdMetric)

Dr. Ellen Forte (CEO and Chief Scientist, edCount)

Dr. Susan Loomis (Independent Consultant)

Dr. Marianne Perie (President, Measurement in Practice)

Dr. Mark Reckase (University Distinguished Professor Emeritus, Michigan State University)

Dr. Lauress Wise (Principal Scientist, Human Resources Research Organization)

The TAC is scheduled to meet for more than 100 hours (approximately 4 hours per month, with additional meeting time following the pilot and operational meetings) to provide technical advice on all aspects of the project to review and revise the mathematics and reading ALDs; this is intended to help ensure that all procedures, materials, and reports are carried out in accordance with current best practices, providing additional validity evidence for the process and results. In addition to frequent meetings and reviews of materials, two TAC members will attend the pilot and operational meetings to observe and provide feedback on the process.

Project Update (May 2021)

The COSDAM meeting on December 7, 2020 included a discussion of the proposed study design and plans for recruiting panelists and conducting the panel meetings virtually given the infeasibility of convening in-person meetings during early-to-mid 2021 in the midst of the COVID-19 pandemic. Committee members asked questions about and emphasized the

importance of protecting secure items in a virtual setting. Moyer explained that plans for maintaining item security were being documented and included the following safeguards: having panelists sign the NAEP non-disclosure agreement (which notes the severe penalties for violations) and repeatedly referring to it; providing Pearson laptops that are locked down and cannot be used for printing documents; using a secure Pearson server with high security protocols; setting up the standard setting platform with a single logon; ensuring that panelists cannot access secure materials outside of the scheduled meeting times; and visually monitoring panelists via Zoom video while they are working with secure materials.

The statement of work for this contract that the Governing Board issued on July 6, 2020 stated that the pilot study shall take place no later than February 2021 (to report results to COSDAM by March 2021) and that the operational study take place no later than early May 2021 (to hold a focused briefing session with COSDAM by the end of May 2021). This timeline was driven by the need for Board action in August 2021 in order to use the ALDs in reporting results for the NAEP 2021 Reading and Mathematics assessments for grades 4 and 8.

On December 27, 2020, Congress passed the Consolidated Appropriations Act of 2021, which rescheduled the mandated NAEP Reading and Mathematics assessments from 2021 to 2022. Consequently, the Board action to adopt reporting ALDs to be used for the release of these results is not needed until August 2022 rather than August 2021 as initially planned. Although there are many planned safeguards for protecting secure materials, there are more limitations inherent in a virtual environment. The recent change to when study results are needed for reporting the next administration means that Board action on the ALDs could be delayed from August 2021 to August 2022 to allow for the increased likelihood of conducting in-person meetings in late 2021 and early 2022.

In conjunction with the March 2021 COSDAM meeting, there was a brief project update indicating that there were plans to modify the project schedule to account for conducting the panel meetings in person in late 2021 and early 2022. A contract modification was executed on March 25 to implement these changes. With this revised schedule, the pilot meeting is being planned for October 25-28, 2021 and the operational meeting is being planned for February 22-25, 2022. Both of these meetings are intended to be in-person, with two TAC members attending to observe the process. To facilitate holding the meetings in-person, the project schedule for reviewing and revising the NAEP Reading and Mathematics ALDs has been extended, so the final ALDs will be presented for Board discussion at the May 2022 Board meeting and Board action at the August 2022 Board meeting. The intention is for the ALDs from this project to be used in the reporting of NAEP results in fall 2022.

The extended schedule also provided an opportunity for the project director, with the involvement of the TAC, to review multiple methods for anchoring assessment items to the achievement levels. The item anchoring to achievement levels is an important aspect of the ALD review process, given that the panelists' review of items associated with an achievement level will inform their statements about what students classified into each level actually know and can do. Several variations of methods for anchoring items to achievement levels were investigated and reviewed with the TAC, including the method used during the 2009 NAEP ALD review

studies, the anchoring method used for the Trends in International Math and Science Study (TIMSS), and anchoring methods using Item Response Theory (IRT) item mapping. The TAC recommended maintaining consistency with previous NAEP anchoring studies (as was also requested by the Governing Board in the Achievement Levels Work Plan) but developing a clear statement of what inferences can be made based on the methodology used.

The review of the anchoring method also included a reevaluation of the criteria for removing items based on item discrimination. In this context, item discrimination is used to differentiate between students classified into adjacent achievement levels. Various discrimination criteria were reviewed with the TAC to determine which process would be best for the study. The TAC recommended not using discrimination criteria even though it was used in previous NAEP anchoring studies because very few items would be removed anyway, and it is possible that these items may be needed for content reasons. That is, if items are dropped from reporting categories that already have a small number of items anchored to them, that will reduce the number of items even further and likely make the task more difficult for panelists.

The Design Document (attached) has been updated to reflect recent changes to the project. The document is very similar to the version reviewed by COSDAM in December 2020; primary changes are highlighted below:

- The panel meetings are intended to take place in-person rather than virtually per the March 2021 contract modification
- The discrimination criteria for determining anchor sets has been removed (page 22)
- At the recommendation of the TAC, the items for Reading will be reviewed by passage, with the items in the passage set being sorted by achievement level (pages 22 and 25)

Next Steps

Over the next few months, project staff will continue working to develop meeting materials and presentations; identify meeting space to hold the panel meetings[3]; and begin panelist recruitment.

[3] Project staff are working closely with Governing Board staff to plan the in-person meetings in compliance with current guidelines from the Centers for Disease Control and the Department of Education's Conference Approval group.

National Assessment Governing Board

Developing Achievement Level Descriptions for Mathematics and Reading

Submitted: April 2021

NAEP ALD Development Design Document

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

This study was funded by the
National Assessment Governing Board under
Contract 91995920R0004.

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

Table of Contents

Executive Summary

The National Assessment of Educational Progress (NAEP), known as the "The Nation's Report Card," provides information on what students in the United States know and can do in various subject areas. As part of its legislative mandate for overseeing and setting policy for NAEP, the National Assessment Governing Board (Governing Board) develops achievement levels that further define expectations of what students should know and be able to do. Achievement on all NAEP assessments is reported using the following achievement levels, in accordance with the Board policy on Developing Student Achievement Levels for NAEP, which are defined as follows:

NAEP Basic – This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for performance at the NAEP Proficient level.

NAEP Proficient – This level represents solid academic performance for each NAEP assessment. 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.

NAEP Advanced – This level signifies superior performance beyond NAEP Proficient.

The policy ALDs apply to all NAEP assessments, regardless of subject and grade. In addition to these achievement levels, content-specific achievement level descriptions (ALDs) are developed to define the expected knowledge and skills for student at each achievement level.

The National Academies of Sciences, Engineering, and Medicine completed an evaluation of NAEP achievement levels in November 2016, which included seven recommendations. Two of the recommendations from the report – recommendations #1 and #3 – were related to NAEP ALDs.

Recommendation #1: Alignment among the framework, 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.

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.

In response to this evaluation of the NAEP achievement levels, the Governing Board developed and approved an Achievement Levels Work Plan in March 2020. To address the recommendations regarding the ALDs from the evaluation, the Governing Board issued a contract to Pearson in September 2020 to develop and conduct anchoring studies using NAEP 2019 data to review and revise the ALDs for grades 4, 8, and 12.

On behalf of the Governing Board, Pearson has developed this Design Document, which describes in detail the activities for the anchoring studies for grades 4, 8, and 12 NAEP mathematics and reading. This document is intended to provide the foundation for all ALD alignment review and revision activities. The Design Document will guide all aspects of the ALD review and revision process, including (1) a set of judgments about the alignment between the policy descriptions and current ALDs and expectations of what students should know and be able to do, based on the items that anchor to each achievement level for each assessment and (2) a set of revised ALDs that could be used as reporting ALDs, describing what students actually know and can do.

For the ALD review and revision studies, Pearson plans to use a model-based approach for reviewing the alignment of the ALDs for NAEP mathematics and reading. The Board's Achievement Levels Work Plan indicated that the methodology for these studies should be similar to previous ALD development and review meetings held in 2009 (Donahue, Pitoniak, & Beaulieu, 2010; Pitoniak, Dion, & Garber, 2010) to reduce the potential for possible inconsistencies from the use of different methods.

The model-based approach includes three stages. The first stage involves conducting statistical analysis to determine the items from the subject and grade that are anchored to each achievement level.

The second stage relies on panels of content experts for each individual assessment. Each panelist will review the items that are anchored to each achievement level to identify the knowledge and skills needed to respond to the items associated with a level. The panelists will then work together to develop common summary descriptions of the knowledge and skills that students in each level know and can do, based on the review of all items anchored to an achievement level. In the third stage, the panelists complete several alignment judgment rounds, comparing the current ALDs for the respective assessment with the summary descriptions. During the alignment judgment rounds, the panelists rate the degree of alignment and take notes regarding the similarities and differences between the ALDs and summary descriptions. The goal of this process is for the panel to make a recommendation regarding whether the current ALDs accurately describe what students in each level are expected to know and be able to demonstrate or if revisions to the current ALDs are needed not only to improve alignment but also to more accurately represent what students within each achievement level know and are able to do.

To facilitate this process, Pearson will use computers during both the pilot and operational ALD review meetings. Using computers along with the online interface in the Pearson Standard Setting website will increase the efficiency of the activities the panelists will need to complete for this study.

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

Section 1: *Achievement Level Descriptions Review Panels* describes the panelist identification and recruitment plan designed to obtain broadly representative and well-qualified panelist groups for all studies.

Section 2: *Briefing Materials* describes the briefing materials sent to panelist prior to each panel study in the ALD review process (pilot study and operational ALD review meeting).

Section 3: *Pilot Study* describes the pilot study designed to incorporate the exact procedures planned for the operational ALD review meeting.

Section 4: *Achievement Level Descriptions Review Tasks and Procedures* describes the ALD review tasks, the nature of the tasks, and the procedures to be implemented prior to and as a part of the

operational ALD review meeting—including how panelists are trained and supported in implementing all activities.

Section 5: *External Feedback* describes the process of obtaining feedback on the results and recommendations from the operational ALD review meeting.

Section 1: Achievement Level Descriptions Review Panels

Pearson will implement a multi-step panelist recruitment plan for the pilot study and operational ALD review meeting. The objective of the recruitment plan is to produce well-qualified panels with a high level of content expertise, consisting of classroom teachers and content experts who will be in the best position to provide the necessary judgments for the respective subject and grade. Panelists for the pilot study and operational meeting will be recruited from across the nation. For each meeting, there will be six panels convened, with each panel focused on a single subject and grade.

Grade 4 Mathematics

Grade 8 Mathematics

Grade 12 Mathematics

Grade 4 Reading

Grade 8 Reading

Grade 12 Reading

A maximum of 48 panelists will be recruited for the operational study, with up to eight panelists recruited for each panel. Additionally, a maximum of 48 panelists with similar background distribution will be recruited for the pilot study, with up to eight panelists recruited for each panel. Panelists in each panel will be assigned to one of two replicate groups, allowing for a comparison of results across groups.

Panels for the ALD alignment review and revision operational and pilot studies will reflect an overall balance of gender, race/ethnicity, geographic location, and urbanicity, i.e., no more than 75% homogenous. Classroom teachers currently engaged in instruction in the respective grade and subject area will compose at least half of the panelists in each panel. At least two of the panelists recruited will be non-classroom educators with curriculum experience within the respective subject, such as state or local curriculum coordinators or higher-education faculty teaching education courses associated with the respective subject and level.

Our goal for this study is to obtain panelists with a high level of content expertise (though not direct experience working on NAEP ALD, item, or framework development) that would enable them to engage in the ALD review process and provide meaningful recommendations and judgments. Pearson will work with staff from the Governing Board along with allied organizations in recruiting panelists.

Overview of Panelist Recruiting Process

A multi-phase process will be used to identify panelists:

Phase 1: Identify nominators through allied organization and state departments of education; contact nominators and ask them to nominate outstanding classroom and non-classroom educators using an online nomination form (e.g., name, contact information, and basic qualifications). Nominators will be asked to briefly describe the rationale for that judgment when provided.

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

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

Prior to finalizing the selection of candidates to participate in the studies, the list of prospective panelists will be prepared and presented to the Governing Board for review and approval.

To optimize recruitment, email and phone calls will be used to communicate with prospective panelists. An honorarium will be paid to panelists for the ALD alignment review and revision operational and pilot studies. Substitute teacher costs will be reimbursed directly to schools based on actual school costs for substitute teacher payments.

Identification of Panelist Nominators

Panelist nominators will be recruited using multiple sources. One source for nominators will be professional organizations that have a strong background in providing professional development in mathematics education or reading and literacy education. Focused will also be placed on professional organizations for minorities. Professional mathematics education organizations that should be used to recruit mathematics panelists should include:

- National Council of Teachers of Mathematics (NCTM)
- National Council of Supervisors of Mathematics (NCSM)
- Association of Mathematics Teacher Educators (AMTE)

Professional English language arts reading, and literacy education organizations that should be used to recruit reading panelists should include:

- National Council of Teachers of English (NCTE)
- International Literacy Association (ILA) (formerly the International Reading Association)
- American Literacy Corporation (ALC)

In addition to these organizations, state education organizations in mathematics or reading and literacy, state superintendents and departments of education, school board presidents, and district and school administrators of public and private education entities will be contacted in the four NAEP regions to propose qualified nominators across both panelist type (classroom teacher and non-classroom teacher educator).

Based on previous experience in recruiting NAEP panelists for achievement level setting meetings, Pearson estimates that 20 percent of the nominators will respond by submitting at least one nominee for consideration. Pearson further estimates that no more than 20 percent of the nominees would meet the qualifications, satisfy the requirements for representation, and agree to serve on the panel. For the pilot and operational studies, an estimated 2400 nominators must be identified to yield at least 480 active nominators, resulting in at least 480 nominees. Assuming that 20 percent of those

nominees will be eligible, meet the distribution requirements for representation on the panels, and be available and agree to serve as panelists, the yield would be the target of 96 panelists, with a target of 48 panelists for the pilot study and 48 panelists for the operational meeting. Pearson will supplement the number of nominators, as needed, to attain the panelists targets.

Prior to the beginning of identifying panelist nominators for the pilot meeting, the project director will meet with the COR and the Pearson meetings team to evaluate the current environment with regard to hosting in-person meetings and determine the feasibility of doing so. This will be reviewed prior to this step to ensure that communications to panelist nominators are accurate concerning in-person meetings. If it is determined, at this point, that in-person meetings will not be feasible, discussions will occur about other options prior to starting recruitment of panelist nominators.

Selection of Panelists

Nominees will be asked to complete an online questionnaire regarding their qualifications and experiences for serving on the panel. Candidates that present the credentials required will be contacted by phone to collect any missing information, verify the information provided, and confirm their willingness to serve on the panel, if selected. The goal is to select the most qualified panelists who are knowledgeable about the related subject at the appropriate level, while maintaining the goal of recruiting a mix of classroom teachers and non-classroom teacher educators for each panel.

Panelists nominated in each panel must meet the following minimal qualifications.

Classroom Teacher Qualifications:

The nominee must meet all of the following qualifications:

- At least five years of overall teaching experience
- At least two years of experience teaching the related subject and the grade
- Judged to be “outstanding” in their professional performance by a nominator

Non-Classroom Teacher Educator Qualifications:

The nominee must meet one of the following qualifications:

- Non-teacher educational staff within school or district with education experience in the related subject and grade
- Curriculum director or content specialist serving schools at the related level or state department of education with education experience in the related subject and grade
- Postsecondary teacher education faculty teaching courses in the related subject and level

The credentials of panelists will be evaluated based on the number and importance of the credentials that are presented. Nominees having no distinguishing credentials will be scored low. Nominees having extensive credentials, including having been named outstanding teacher/teacher of the year and/or being actively engaged at the state or national level in professional activities within the specific subject and level, will score high. The scoring scheme differs for each panelist type (classroom teacher and non-classroom teacher educator). Nominees with the highest scores are given top priority by being considered the best-qualified candidates and being placed at the beginning of the candidate list. The selection process then selects panelists to reach the targets for representation listed above, with nominees having the highest qualifications being the first selected each time. All panels will be selected to have approximately equal proportions of males and females and equal proportions of representation from each of the four NAEP geographic regions. Every attempt will be made to create panels in which at least 25 percent of the panelists self-identify as a minority.

Prior to the beginning of both recruiting panelists and the final selection and notification of panelists, the project director will meet with the COR and the Pearson meetings team to evaluate the current environment with regard to hosting in-person meetings and determine the feasibility of doing so. This will be reviewed prior to this step to ensure that communications to panelists are accurate concerning in-person meetings. If it is determined, at either point, that in-person meetings will not be feasible, discussions will occur about other options prior to starting recruing panelist.

Each panelist for the pilot study and operational ALD review and revision meeting will be given an honorarium. Pearson acknowledges that the funds available to offer panelists are not commensurate with their contribution. They will emphasize that panelists' participation in the ALD review and revision project represents an exceptional contribution to education in the United States.

Section 2: Briefing Materials

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

- Confidentiality agreement
- Purpose and overview of the meeting
- Meeting roles and responsibilities
- Request for reimbursement form
- Meeting agenda
- NAEP framework for the relevant assessment
- Other materials identified as appropriate

Communication with panelists will encourage them to engage with the briefing materials as those materials are intended to serve as a foundation for successfully carrying out the process designed for each panel. Pearson staff will be able to monitor panelist activities on the website to determine which materials a panelist has accessed.

Section 3: Pilot Study

In late October 2021, Pearson will conduct a pilot study of 4 days to implement the exact meeting procedure planned for the operational ALD alignment review and revision meeting. Conducting the pilot study at this time offers an opportunity to preview, revise, and resolve issues prior to the operational ALD alignment review and revision meeting in February 2022. To maintain uniformity of conditions, the pilot study and operational ALD alignment review and revision meetings will be held using the same process, including the same agenda of activities, and all steps in the pilot study will be the same as those planned for the operational ALD alignment review and revision meeting. By fully replicating the process of the operational study, the pilot will provide the information needed to determine whether any modifications are needed for the operational study.

The Technical Advisory Committee (TAC) will be consulted for advice and recommendations regarding details of the design prior to and after the pilot study. Up to two members of the TAC will be asked to observe the pilot study. All pilot and operational study materials will be based on data from the 2019 operational administration of the NAEP mathematics and reading assessments. The pilot study has the following goals:

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

Given that all steps in the pilot study represent those planned for the operational ALD alignment review meeting, details on the process for both the pilot study and the operational meeting are provided under Section 4 of this document which clearly describes the operational ALD alignment review meeting.

A pilot study report will be prepared no later than December 2021, for presentation to COSDAM

during a webinar. The timing of the pilot study allows for preparation of the report and review of the report by the COR, COSDAM and TAC so that improvements to the process can be made in advance of the operational ALD review meeting.

Section 4: ALD Alignment Review Tasks and Procedures

Pearson will conduct an operational ALD alignment review meeting in February 2022. To maintain uniformity of conditions, the operational ALD alignment review meeting will be held with the same agenda of activities used for the pilot study, with the exception of adjustments made based on improvement to the pilot study process. This section describes the ALD alignment review procedures and tasks that Pearson will implement during both the pilot study and operational ALD alignment review meeting and includes information about the configuration of panels and materials, training of panelists, the collection of panelists' ratings, and the feedback given to panelists.

The current plan is to conduct both the pilot and operational meetings in-person. At this time, it is believed that an in-person meeting can be held safely and successfully, implementing sanitization and social distancing where possible. The CDC has prioritized educators in the vaccination schedule and it is expected that this will not negatively impact recruitment. However, the pandemic is being carefully monitored, and contingency plans are being developed.

The operational meeting will involve 48 panelists, with approximately eight per subject/grade. The panelists within each subject/grade will be assigned to two groups with approximately equal representation, which will be used as replicate panels during the meeting, since they will complete the same process receiving training and modeling from the same facilitator, however they group facilitation will be different. The TAC will be consulted prior to the operational meeting, and up to two members of the TAC will be asked to observe the meeting. All operational materials will be based on data from the most recent 2019 administration of the grade 4, grade 8, and grade 12 NAEP mathematics and reading assessments.

As previously indicated, Pearson will implement a model-based anchor approach and use the Pearson Standard Setting website platform to facilitate key aspects of the ALD alignment review process, including panelist training, review of ALDs, housing and reviewing of selected anchor item

sets, recording of panelists' feedback and ratings during each round of review, provision of feedback, and evaluation of the ALD alignment review process. The Pearson Standard Setting website will be set up to guide panelists through the steps of the ALD review process, with facilitators having the ability to restrict or provide panelists with access to sections of the site or activities, as needed. The night before the first day of the pilot study and the operational meeting, Pearson will reset the panelists' passwords for all panelists to a common meeting password. Additionally, access to the sites is controlled by Pearson administrators. The sites are deactivated overnight and during any extended breaks during the day.

The model-based anchored approach includes three stages. The first stage will involve conducting statistical analyses to determine the items that anchor to each achievement level (*NAEP Basic, NAEP Proficient,* and *NAEP Advanced*). Additionally, an anchor set will be developed for items that map to the higher end of the region below *NAEP Basic*. The second stage relies on panels of content experts for each individual assessment. The panelists in each committee individually review the items that are anchored to each performance level. While reviewing they items, they will write a description of the knowledge/skills required to answer the item correctly or receive a specific score point. It is from those individual item descriptions that the group level summary descriptions are developed for each achievement level. In replicate panels, they create summary descriptions of what students in each level know and can do based on the body of knowledge and skills measured by the items. In the final stage, the panelists compare the current ALDs for the respective assessment with the summary descriptions, noting the similarities and differences, to make a recommendation regarding whether the current ALDs describe what students in each level know and can do or if revisions to the current ALDs are needed to improve alignment. If it is determined that revisions are recommended, the panel provides recommended modifications to the ALDs.

Computer Use and Software

Pearson will use computers during the pilot study and the operational ALD review meeting. The use of computers and an online interface will reduce the time required for panelists to complete most steps in the ALD review activities. In addition, the use of computers will allow the panelists to interact with the items as students did, such that panelists are better able to understand what examinees would have to know or be able to do.

The computer will be a laptop computer (provided by Pearson) that is used by panelists to access the online Pearson standard setting website interface. With the review of materials required during the meeting, the panelists will also be provided an external monitor, which will provide them more screen area during the meeting to complete activities. As a part of supporting both efficiency and accuracy of the ALD review process, Pearson will use computers with access to the Pearson Standard Setting website, with a section specifically designed for the NAEP ALD review meeting. The NAEP website can be accessed only through a user identification code (ID) assigned by Pearson. Permissions will be set up for each user ID so each panelist can access only the materials he/she will be using.

The use of the website will reduce the time required for panelists to complete multiple steps in the process, since they will be able to access materials and complete activities within the website interface. As a website-based system, all materials and data will be stored, organized, and accessed through the website, which will ease the demands of development of materials and ensure the consistency of the materials with which the panelists interact. The use of the website will also ensure the security of the materials during the standard setting meeting, since the system will be used throughout the ALD review process and requires a secure login by all users. Pearson designed the interface for the NAEP ALD review process to have the following features:

- Simultaneous access by multiple users, with each individual user assigned a profile which defines their level of access to the site, including the materials the user can access and the specific functionalities available. Facilitators and site administrators can use conditional access

features based on date, activity completion, or custom settings to define access to materials and functionalities of the site.

- User access prior to the ALD review meeting, so users can interact with materials and experience training that will prepare them for the meeting and the ALD review process.
- Consistent user interface throughout the entire ALD review process, from pre-meeting work to post-meeting feedback.
- Management of storage and access to materials through the website, as a single access point for the meeting.
- Assignment of panelists to materials and items, based on panelists grouping.
- Embedded quality control features through conditional settings within activity that limit the range of responses to valid entries only and ensure no blank entries are accepted.
- Export of panelists' item review notes, as a group or individually. Facilitators will be able to view the panelists' notes within the website, to verify panelists' responses.
- Secure storage of all panelist information and judgments within the website, with access restricted to facilitators and site administrators.

Preparing for the ALD Alignment Review Meeting

This section describes key activities Pearson will complete before the pilot and operational ALD alignment review meetings that contribute to the success of the meetings.

Development of Anchor Item Sets.

Prior to selecting the proposed anchoring approach outlined in this section, several variations were investigated to determine the effect of using different approaches, including the anchoring method used during the 2009 NAEP anchoring studies, item response theory item mapping approaches, and the anchoring method used for the Trends in International Math and Science Study (TIMSS). The investigated methods and the data were shared and discussed with the TAC. Each method resulted in different anchor item sets, which resulted from different conceptualizations of the set of students used to calculate the conditional probability utilized to anchor the item to an achievement level. It was

recommended by the TAC that the methodology from the previous alignment studies for mathematics and reading in 2009 (and specified by the Governing Board in the Achievement Levels Work Plan) would be the most defensible approach. This is because of the need to make consistent inferences about what the ALDs represent across the years by using the same anchoring methodology. That is, the anchor sets across years are created in the same manner and therefore the descriptions of the achievement levels represent the same range of students.

The development of the anchor item sets starts by grouping performances representing individual students from the most recent (2019) administration of the grade 4, grade 8, and grade 12 NAEP mathematics and reading assessments into achievement levels. The achievement level classification for each student is based on the average of their NAEP "plausible values" and the relationship to the boundaries of the achievement levels for the respective assessment. A student will be classified into either *NAEP Basic*, *NAEP Proficient*, or *NAEP Advanced* if their mean plausible value is at greater than or equal to the cut score for the respective achievement level. A student is classified into the region just below *NAEP Basic* when their average plausible value is below the cut score for *NAEP Basic* and the region between the mean plausible value and two standard errors above the mean includes the cut score for *NAEP Basic*. This approach will use all students in the NAEP sample from the most recent administration in 2019 to ensure that there are sufficient students associated with each achievement level for the analysis to determine each anchor item set. This is an approach has been utilized in previous NAEP anchor studies.

After performance indicators for students are assigned to an achievement level, the conditional *p*-value, or probability of each student in that achievement level answering each item correctly, will be calculated using the IRT statistics from the most recent administration of the assessments. The conditional *p*-value for students across a given level will be averaged to derive the anchoring probability for that item or score point for multi-point items. Each item or score point will be assigned four

conditional *p*-values, one each for below *NAEP Basic, NAEP Basic, NAEP Proficient,* and *NAEP Advanced*, which represent the average performance on the item of the typical student within the four achievement levels. Items will be anchored to the first achievement level where the conditional *p*-values for the achievement level are greater than or equal to 0.67.

In 2009, in addition to the conditional *p*-values, item discrimination values were calculated for each item and achievement level and used as a determinant in the final anchor sets. Analysis of various discrimination criteria were reviewed with the TAC, including the initial criateria of the 40th percentile, a standard deviation criteria, a fixed value criteria, and using no criteria. Preliminary analysis across the different methods indicated that the discrimination criteria was removing very few items. In some grades of mathematics, however, the content areas had a limited number of items that anchored to them and dropping any item could risk adequate content coverage. The minimal number of items dropped due to the different discrimination criteria and the possible impact on the content coverage of the items resulted in the recommendation to not utilize a discrimination criteria for this study. After discussion with the TAC, it was decided that a discrimination criteria would not be used as part of the anchoring process.

Based on the anchoring criteria, items will be classified into one of five categories: (1) just below the *NAEP Basic* level, (2) *NAEP Basic* level, (3) *NAEP Proficient* level, (4) *NAEP Advanced* level, or (5) does not anchor. The items in the anchor item sets for the respective assessment will be grouped by content area. By reviewing the items within a content area, across all achievement levels, the panelists will be able to maintain a consistent focus on the knowledge and skills associated with the content area. For mathematics, the items associated with a content area will be ordered by achievement level from below the *NAEP Basic* level to the *NAEP Basic* level, to the *NAEP Proficient* level, and then finally the *NAEP Advanced* level. Within an achievement level, the items are in decreasing order of conditional *p*-value, so the easiest item associated with the achievement level is first and the most difficult item is last. In this way, panelists will see a progression in what students know and are able to demonstrate while

working through the items that anchor to that achievement level. For the reading assessment, the items within a content area will be presented as sets with their associated passage. The passages will be ordered by average *p*-value. The items withing a passage set will also be ordered by achievement level and by average conditional *p*-value.

Division of Panelists into Replicate Panels. In order to assess the degree of internal validity, two replicate panels will be created within a panel. Approximately three to four panelists will be in each replicate panel. Pearson will assign panelists to a replicate panel with the intent of creating panels that are as equivalent as possible to one another. The purpose of the replicate panel is to be able to assess the degree of internal validity. The summary statistics from the replicate panel rounds of individual comparisons will be calculated and compared. In addition, a qualitative analysis will be conducted to assess the degree to which the summary descriptions created by each replicate panel vary.

Provision of Advanced Briefing Materials. As discussed in Section 3 Briefing Materials, panelists will have access to advanced materials through the Pearson Standard Setting website. Pearson will send each confirmed panelist access to the materials for their review and familiarization prior to the operational meeting, including information to log into the website and change the assigned password. The first time the panelist logs in to the website, he or she must read and electronically sign a nondisclosure agreement. Once signed, the panelist will be guided through a brief online training for using the website before having access to the non-secure advanced materials. Panelists will then use an online checklist to progress through the review and to ensure that each document is reviewed. Panelists will also have access to an orientation activity that includes an overview video as well as multiple engagement check-ins to gauge interaction and preparedness.

Training of Facilitators. The ALD alignment review and revision study will involve two content facilitators as well as a process facilitator per panel. There are six panels for each study, three

reading panels and three math panels. The content facilitators are selected for their expertise and experience in instruction in the related subject and grade as the framework they bring. The process facilitator is selected for the expertise and experience conducting meetings they have. Recognizing that facilitators may introduce individual difference that can result in slightly different instructions, content and process facilitators will be properly trained to implement the process uniformly and as intended. Pearson will prepare the PowerPoint presentations that facilitators will use during the meeting. In addition, facilitator handbooks will include the tables and graphs, a script for providing instructions, a description of the activities and an explanation of the feedback. Facilitators will attend a one-day virtual training prior to both the pilot study and the operational meeting. The project director overseeing the activities will lead the training. In addition, the facilitators and project director will do a walkthrough of the entire meeting the day before the pilot study and the day before the operational meeting.

Preparation of Pearson Laptops. Pearson will be providing each of the panelists a Pearson laptop that has been configured to have the appropriate software needed to access the NAEP items as well as the Pearson Standard Setting website. The laptops will be shipped to the meeting site and stored in a secure location when not in use. Additionally, each panelist will be provided an external monitor to ensure that they have the screen area to complete their work with minimal need to switch tabs or screens during the process.

Preparing Panelists for the ALD Alignment Review Process

Provide an Orientation. The operational meeting will have 48 panelists, with approximately eight panelists assigned to each panel. The six panels, along with the process and content facilitators, will begin the meeting with introductions and a description of the panelist recruitment process.

Provide an Overview of the Alignment Review and Revision Process. Facilitators will provide an overview of the purpose of the ALD alignment review and revision in general and description of

the process that will be used. Panelists will receive training in the key components of the NAEP framework for the related assessment and the rationale supporting these components. The administration and sampling used for NAEP administration will also be presented. The process facilitator will describe the process that was used to create the anchor item set and an orientation to the information that is provided for each item. The panelists will also be trained in how the items are scored with the item key or scoring rubrics. We will also orient the panelists in the structure of the alignment judgment process, including the anchoring of items to specific achievement levels and the ordering of the items within achievement levels.

Individual Item Review and Summary Descriptions

Item Review. During the individual item review process, panelists will review all items within the anchor item set. Prior to panelists beginning the item review process, the facilitator will demonstrate how to use the website to review each item, view item information, and collect individual panelists' notes about each item using a practice anchor set. Additionally, the content facilitator will model to the panel how to approach the items in the anchor set and the process of developing item descriptions. The facilitator will also show how panelists how to convert the item description into a summary statement about what students know and can do for an achievement level. The panelists will work independently to review each item and create a description of the knowledge and skills demonstrated by students who answered the item correctly or who received a particular score for constructed-response items. During this step, panelists will review items within the same content area. For mathematics, items will be grouped by increasing achievement level (including the items anchored to just below the *NAEP Basic* level) and conditional *p*-value, from easiest to hardest. For reading, the passages will be ordered by increasing difficulty (from easiest to most difficult, based on average conditional *p*-value) and all items within a passage set will be presented to panelists before they review the next passage. Items within a passage will be ordered by increasing achievement level and conditional *p*-value. When

panelists have completed their review of the anchor items for an achievement level, they will be asked to write a single summary statement about what students within the three NAEP achievement levels (NAEP Basic, NAEP Proficient, and NAEP Advanced) know and can do, based on the item descriptions. Once panelists have completed their review of items across all achievement levels, they will be asked to consider what students in one achievement level can do compared to the adjacent achievement level(s). Panelists will complete the entire process for a content area before progressing to the next content area.

Summary Descriptions. After individually writing descriptions for each item for a content area, the panelists will convene in their replicate panel to review the collected set of knowledge and skills associated with the items within an achievement level and construct a summary of what students performing in that content area and achievement level know and can do. (For reading, the panelist's item descriptions will be sorted by achievement level and conditional *p*-value so that they can reference the items during the discussions). Prior to beginning the replicate panel development of summary descriptions, the facilitator will model the process of using the item descriptions to develop summary descriptions by achievement level. This process will start with the *NAEP Basic* level, then the *NAEP Proficient* level, then the *NAEP Advanced* level. They will create their summary descriptions for each achievement level, within a content area, in the website using a shared document. The summary descriptions will be captured by the content facilitator assigned to that replicate panel. Once the replicate panel has completed the summary descriptions for a content area, they will then begin individual work for the next content area. After each content area of individual review, they will meet in the replicate panels to create the summary descriptions for that content area. When all content areas are completed, the panelists will progress to the alignment judgment rounds with replicate panel discussions, whole panel discussions, and ALDs revisions, if needed.

Alignment Judgment Rounds with Panel Discussions

The process of creating and reviewing ALDs is an iterative process, where individuals provide individual alignment judgments, followed by replicate panel discussions and whole panel discussions that are informed by results from the alignment judgments, resulting in more informed judgments in subsequent rounds. There will be three alignment judgment rounds. Prior to the first alignment judgment round, the content facilitator will provide the panelists training about the meaning of alignment between the summary statements and the policy ALDs as well as the current content ALDs.

Alignment Judgment Rounds. The replicate panels will be a significant part of the review and judgment process. After the panelists complete their individual item review the panelists will then meet in their assigned replicate panels to discuss their individual summary statements and to develop a replicate panel set of summary statements for that content area. They will toggle between individual review of items by content area and replicate panel development of summary descriptions until all sub-content areas have been addressed. To identify areas of alignment and lack of alignment, panelists will then work independently to make comparisons between the replicate panel summary descriptions with the policy descriptions and current ALDs. Panelists will be asked to assess the degree of alignment of the replicate panel summary descriptions by content area and achievement level to the policy descriptions and to the ALDs. If they indicate weak to moderate alignment, they will be prompted to provide suggestions/comments for why the alignment is weak and what modifications could be made. Additionally, they will make a judgment on the overall alignment of the summary descriptions as a whole to the policy descriptions and current ALDs.

After the first round of individual alignment judgments, panelists will meet back in their replicate panels and discuss why they made their specific alignment judgments. If necessary, they can make adjustments to the summary descriptions to clarify or modify the language to better indicate the original intended meaning. The panelists will then proceed with another round of individual alignment

judgments. They will then meet back as a whole panel and review summary statistics of the alignment judgments made and have a whole panel discussion similar to the replicate panel discussion. Where necessary, they will visit the summary descriptions to evaluate if the difference in the judgments relates to differences in the summary descriptions between the replicate panels. Panelists will be shown the summary descriptions from the two replicate panels side-by-side to evaluate the degree to which they are similar or different and develop a set of whole panel summary descriptions. The panelists will have one more opportunity to make individual alignment judgments based on the whole panel summary descriptions.

After the third round of individual alignment judgments, a whole panel discussion will take place. For the final whole panel discussion, panelists will be asked to focus on any content areas that still indicate greater than 50% panelist agreement of weak alignment or greater than 66% panelist agreement of moderate or weak agreement with the current ALDs. The panelist will review the ALDs for these subcategories and have the opportunity to suggest edits to the ALDs to improve alignment.

Cross-grade Review. Before the end of the meeting, the panelists will be brought together to review the reporting ALDs developed by each group, to review the terminology used across the achievement levels and grades to ensure that they clearly delineate progression of skills across grades and levels. This discussion will be led by one of the process and content facilitator teams. Any suggested revisions to the reporting ALDs will be determined by the panelists that constructed them, since they are most familiar with the content and discussions, but recommended changes may be provided by any panelists. The recommended ALDs will be the result of this final meeting.

Process Evaluations. Procedural evidence refers to the appropriateness of the procedures and how well those procedures were implemented. Evidence for procedural validity may come from a number of sources, including criteria for selecting panelists, the justification for the method, the quality of the implementation of the procedure, and the completeness of the documentation of the process. As

another source of evidence of procedural validity, panelists will be asked to complete evaluation forms after each major activity of the process. Evaluations will include both selected-response and open-ended questions that address the panelists understanding of the process and confidence in the results. No key-entry by staff is required because panelists will use the secure website interface to complete their evaluations. Panelist entries will be available for viewing using the facilitator login to the website. Facilitators will scan written responses for possible problems as they are collected during each day. Summary statistics will be computed for all ratings items and written responses. These analyses will be reviewed in real time throughout each day, and any sources of confusion will be identified for clarification with individual panelists or the panel as a whole. The operational meeting must be completed in time for the recommended ALDs to be used for reporting the next NAEP mathematics and reading assessments in grades 4, 8, and 12 during fall 2022.

Section 5: External Feedback

Pearson understands the potential contribution of external feedback to the review and revision of ALDs for the NAEP mathematics and reading assessments. Pearson has designed a strategy for collecting and using feedback from content experts and NAEP stakeholders for the panel recommendations for revising the ALDs.

Pearson plans to implement a vigorous and targeted approach to soliciting feedback regarding the panel recommendations through personal contact with key leaders and members of stakeholder groups, and through the use of a simplified and directed format for reviewer response. Pearson will create a website to obtain external feedback on the panel recommendations for revising the ALDs resulting from the operational ALD review meeting. The website will provide a means for stakeholders and the public to find information about the study design and the panel recommendations and to provide feedback. Prior to opening the site to the public, Pearson will submit the site to Governing Board staff for review and approval.

Pearson will solicit comments from content persons, technical persons, and persons in education-related organizations that are known to use NAEP. When organizations are engaged, a personal call from project staff will establish the initial contact and explain the importance of the request. For collection of comments from members of organizations, Pearson will ask the organization leadership to communicate the request for feedback to the membership and to encourage members' cooperation and response. In all cases, a personalized email message will be sent to explain the purpose of the request for comment and to provide a link to the website having background about the project, instructions, and questions for respondents to consider. Reviewer comments can be recorded through the link, downloaded by Pearson staff, and saved for analysis. Follow-up emails will be sent to encourage responses and to contact individuals for clarification and additional information, should that be necessary. The feedback will be summarized and presented to members of the panel during a virtual meeting, to allow the panelists to make any revisions to their recommendations before the final ALDs are presented to the Governing Board.

References

Bourque, M. L., (2009). *A history of NAEP achievement levels: Issues, implementation, and impact, 1989-2009*. Washington, D.C.: National Assessment Governing Board.

Donahue, P., Beaulieu, N., Freund, D., & Pitoniak, M. (2009). *Report on the 2009 Reading Achievement Level and Scale-Anchoring Study (Draft)*. Princeton, NJ: Educational Testing Service.

Egan, K. (2021). *Reviewing achievement levels and cut scores.* Washington D.C.: National Assessment Governing Board.

Loomis, S. (2016). *Anchor Studies for Analysis of NAEP Achievement Levels*. Prepared for presentation to an Expert Panel convened by HumRRO for the National Assessment Governing Board, Arlington, Virginia, on July 12–13, 2018.

Mullis, I.V.S, Cotter, K.E, Centurino, V.A.S, Fisbein, B.G & Liu, J. (2016). Using Scale Anchoring to Interpret the TIMSS 2015 Achievement Scales. In M.O. Martin, I.V.S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 14.1-14.47). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timss.bc.edu/publications/timss/2015-methods/chapter-14.html

Pearson (2016). *Developing achievement levels on the National Assessment of Educational Progress in grade 8 technology and engineering literacy: Technical report*. Iowa City, IA: Author.

Perie, M. (2018). *Validating achievement level descriptors.* Paper prepared for presentation to an Expert Panel convened by HumRRO for the National Assessment Governing Board, Arlington, Virginia, on July 12–13, 2018.

Framework Development Processes

Under the leadership of the Assessment Development Committee (ADC), the Board updated its Framework Development policy in March 2018. One of the primary revisions reflected in the current policy was to account for the process of updating existing frameworks; the previous policy emphasized the development of new frameworks and contained little explicit guidance on monitoring and revising frameworks without starting from scratch.

The current policy has now been in place for three years and has guided the updates of the NAEP Mathematics Framework (adopted by the Board in November 2019) and the NAEP Reading Framework (currently under Board consideration). Leadership of ADC and COSDAM have identified a need to evaluate the extent to which the current policy and procedures are meeting the intended goals and determine whether any aspects need to be revisited.

To support a joint ADC-COSDAM session on this topic, Board staff commissioned two papers:

- As a consultant, former Governing Board Executive Director Cornelia Orr synthesized historical information on NAEP framework development, including:
    - Initial NAEP legislation and how it has evolved in its requirements for framework processes and outcomes
    - Board policy and how it has evolved in its requirements for framework processes and outcomes
    - Policy contexts and professional standards that have shaped framework processes
    - Procedures the Board has used to adhere to law/policies/professional standards
    - Description of how framework procedures have evolved over time
    - Reflections on why framework procedures have evolved the way they have, in light of policy contexts, professional standards, laws, etc.

- As part of the Board's contract for Technical Support in Psychometrics, Assessment Development, and Preparedness for Postsecondary Endeavors, the Center for Assessment (under subcontract to the Human Resources Research Organization) prepared information on how NAEP framework development relates to procedures for developing other assessments, including:
    - Summarizing elements of framework processes for state, national, and international assessments
    - Comparing these framework processes, articulating similarities and differences
    - Listing and describing best practices in framework processes, in general
    - Evaluating which best practices are appropriate for NAEP's legislative mandates, e.g., curricular-neutrality, pedagogical-neutrality, etc.
    - Describing how current NAEP framework processes reflect or do not reflect these NAEP-appropriate best practices

The papers have been completed and will be the focus of a joint ADC-COSDAM meeting that is in the process of being scheduled to occur in June.

HISTORY, POLICY, AND DECISION POINTS

Developing NAEP Assessment Frameworks

April 2021

Cornelia S. Orr, Ph.D.

History, Policy, and Decision Points for Developing NAEP Frameworks

Table of Contents

I. Introduction and Historical Overview

The National Assessment Governing Board (Governing Board) is an independent, bipartisan organization that sets policy for the National Assessment of Educational Progress (NAEP), commonly known as The Nation's Report Card. Since its creation by Congress in 1988, the Governing Board has overseen and set policy for NAEP by identifying subjects to be tested, determining and approving the assessment content, setting achievement levels for each assessment (i.e., NAEP Basic, NAEP Proficient, and NAEP Advanced), improving the reporting of results, and planning and executing initial releases of NAEP Report Cards.

The 26 members of the Governing Board includes governors, state legislators, state and local school officials, educators, researchers, business representatives, and members of the general public, who are appointed by the U.S. Secretary of Education. As part of the Governing Board's policy setting role, it adopts policy statements and resolutions for NAEP which provide guidance about the implementation of NAEP to persons and organizations working with and on behalf of the Governing Board. The Governing Board's policies align with the purpose of NAEP to provide fair and accurate measurement of student academic achievement. Members of the Governing Board and the National Center for Education Statistics (NCES), working in tandem, conduct activities to implement NAEP and communicate NAEP results to diverse audiences.

This paper provides a summary of the history of the Governing Board framework development processes and the evolution of the policy that now governs how the Governing Board determines the content for NAEP. It explains how changes have occurred over time and the implications for current and future framework development. This paper also describes key decision points in this process, for example, when the Board involves external partners and stakeholders in updating or revising frameworks, and describes the Board's role in approving frameworks.

What Is a NAEP Assessment Framework?

In the 2009 publication *A History of NAEP Assessment Frameworks*, Carol Jago provides this definition.

> NAEP frameworks describe the assessment objectives and design for national tests in reading, mathematics, writing, science, history, civics, economics, foreign languages, geography, and the arts. Governing Board policy dictates that these assessments must be valid, reliable, and based on widely accepted professional standards. (Jago, 2009, p. 1.)

NAEP assessment frameworks "are conceptual, overview documents that lay out the basic structure and content of a domain of knowledge and thereby serve as a blueprint for assessment development." (Haertel, et al., 2012, p. 14) Framework documents typically define the content area in two dimensions: (1) the content and skills to be tested, and (2) the cognitive processes and complexity assessed within the content area. Further, the framework specifies

the types of test questions to be used and the balance of content (weighting) to be assessed. More specific details about developing items to measure the content and cognitive processes at differing levels of cognitive complexity are contained in a companion "specifications" document for each framework. NAEP assessment frameworks provide both the "what" and the "how" for NAEP and have been used by the Governing Board since its inception in 1988.

NAEP before the Governing Board

Since the initial administration of the NAEP in 1969, much has changed in the education landscape and the assessment itself. In the early years, the assessment was developed to provide content-specific information useful to educators. The NAEP reports were designed to provide data on the success levels on a task (percent correct) and not an overall score. Summary scores were avoided because there were concerns about federal government intrusion into state and local school district decisions about education. (Lehmann, 2004; Selden, 2004) Similar concern exists today and probably always will.

In 1969, the responsibility for implementing the national assessment was given to the Education Commission of the States (ECS)—an organization of state leaders that could be "trusted" not to infringe on the rights of its members. While this arrangement continued successfully for several years, a 1976 government report issued by the Comptroller General contained a plea to "make NAEP more useful." (U.S. General Accounting Office, 1976) New federal legislation in 1978 brought changes to the oversight and organization of NAEP and established an Assessment Policy Committee of 17 members (the precursor to the Governing Board). In 1982, a major study critical of NAEP was published which said NAEP was underdeveloped and underutilized, and of "apparently negligible influence." (Wirtz & Lapointe, 1982)

In 1986, then Secretary of Education William J. Bennett formed a distinguished group of state leaders, called the Alexander-James study group. The group questioned the narrow range of subjects that NAEP was covering—due mainly to inadequate funding. Their report was reviewed by the National Academy of Education, and their review was incorporated in the report prior to publication. (Alexander & James, 1987) The debate which followed resulted in revised legislation and more changes for NAEP. The 1988 reauthorization of NAEP not only created the National Assessment Governing Board, it gave the Board specific responsibilities in regard to NAEP. One of these responsibilities was determining what would be assessed and how.

Anticipating the 1988 legislation that would permit voluntary state participation in NAEP, the National Assessment Planning Project (NAEP, 1988, pp. 5-6) was established to make recommendations for the 1990 mathematics assessment. The project utilized a process for developing objectives similar to that described in the legislation which authorized NAEP through June 30, 1988. However, it was expanded to ensure careful attention to formal mathematics objectives of states and some local school districts, and to elicit the opinions of practitioners at the state and local level about the content that should be assessed. This

involvement was seen as a key component to encourage the participation of states, particularly given that NAEP would produce state report cards. The effort to identify and review the objectives provided the assurance states wanted about the content being assessed. (Selden 2004, pp. 195-199)

1987-1990 Overlap: NAEP and the Governing Board[1]

The first assessments administered after the 1988 establishment of the Governing Board were in reading and mathematics in 1990. Those assessments utilized the NAEP reading and mathematics objectives being developed in anticipation of the 1988 law. These objectives were developed and reviewed as part of the NAEP National Assessment Planning Project. The 1990 NAEP Mathematics Framework and Reading Framework were published in November 1988 and April 1989, respectively, by ETS on behalf of NAEP. (NAEP, 1988; NAEP, 1989)

The development of the frameworks utilized a consensus development process. The 1988 Mathematics Framework described these elements. (NAEP, 1988, pp. 6-9).

- A seventeen-member Steering Committee included policy makers nominated by national organizations. One member was also on the Mathematics Objectives Committee.
- An eleven-member Mathematics Objectives Committee comprised of a teacher, a school administrator, mathematics education specialists from various states, mathematicians, parents, and citizens recommended objectives for the assessment.
- The draft objectives were distributed to the mathematics supervisor in each of the 50 states and also to 25 mathematics educators and scholars for their review.
- Incorporation of comments and revisions were made by the Mathematics Objectives Committee with the final recommendations approved by the Steering Committee.
- After the objectives were submitted to NCES, they were provided to the Assessment Policy Committee which approved the Project recommendations.[2]

Because NAEP would now produce state report cards, both the reading and mathematics process to develop objectives paid careful attention to the formal objectives of states and to the opinions of practitioners at the state and local level. In particular, efforts were made to integrate new theory and research on the learning and teaching of these subjects and to reflect the innovative approaches of assessments being developed. (NAEP, 1989, p. 7)

The Governing Board Framework Development Policy Overview

Beginning with assessment frameworks adopted for the 1992 assessment, Governing Board staff managed the process of soliciting and engaging contractors, and overseeing the work of

[1] A more detailed presentation of the historical activities related to the history of NAEP and the Governing Board is found in Appendix A.

[2] The Assessment Policy Committee provided policy oversight for NAEP and was established in the 1978 NAEP reauthorization. Also see discussion on page 2 and Appendix A.

committees charged with identifying the content for the assessments. A Governing Board staff member attending the second meeting of the Governing Board observed, "One of the most important issues considered at the January 1989 meeting was developing a 'consensus process' for determining the content of the 1992 reading assessment." (Bourque, 2004, p 205) The development of the framework was to be carried out via a contract with the Council of Chief State School Officers (CCSSO). The CCSSO staff recommended the principles summarized below which were contained in the January 1989 Governing Board meeting materials.

1. The process should be participatory, visionary, iterative, structured, explicit, stable, and supported by adequate resources.
2. The management of consensus committees should be in a value-free way, to encourage opinions and avoid curtailing or intimidating the participants.
3. The process should be mutually educational for those involved.
4. Values and constraints for the process should be stated up front.
5. Changes in the structure or rules of the consensus process during the process must be avoided.
6. Solicitation of comments representing the field is needed only in response to the draft recommendations.
7. Board members must decide carefully with which people they will work.
8. Work on subject-matter objectives, procedural, and analytic plans should be a staff function of the governance process, and review by the field should be part of the process.
9. The consensus process should be self-evaluating.
10. The planning process should have a built-in buffer to ensure that the recommendations are thoughtful and appropriate.

Bourque, the Governing Board Assistant Director for Psychometrics from 1989 to 2001 and an observer of the consensus processes for reading, writing, U.S. history, world geography, science and civics indicated these 10 principles were "in large measure what govern the work of the groups" who make the framework recommendations. (Bourque 2004, p. 206) The CCSSO report at the January 1989 meeting also included the recommendation that the Governing Board develop an explicit policy to direct those developing objectives for NAEP. When one considers the Governing Board workload to adopt frameworks between 1989 and 2002[3], it is not surprising that the explicit policy did not emerge until 2002. It is reassuring that similar practices as those ultimately included in the 2002 Framework Development Policy were in place before they were codified.

In 2018, the Governing Board revised the Framework Development Policy, primarily to add a provision for updating frameworks when a complete framework revision was not needed. The policy had originally been conceived for the development of new frameworks. This revision

[3] The Governing Board adopted the following frameworks between 1989 and 2002: Reading (1990), Writing (1990), Science (1991), U.S. History (1992), Geography (1992), Arts (1994), Civics (1996), Writing (1996), Mathematics (2001), Foreign Language (2000), Economics (2002).

also included streamlining some wording and moving procedural details to the contracting documents called statements of work. Details about these revisions will be discussed in a later section.

II. Legal Requirements for Assessment Frameworks

Are "frameworks" required in the law?

Technically, no. The current and previous versions of the Congressional authorization do not use the term framework. 'Assessment framework' is a construct used to distinguish what will be tested from what is taught (curriculum standards or instructional objectives). Some assessment programs use the term test blueprint or test specifications. While the construct of an assessment framework is not unique to the Governing Board, it is the term that was chosen. The NAEP assessment frameworks do not cover every aspect a content area, especially what students should be taught and how; they simply describe which aspects of the content area will be tested on NAEP and the how that content will be assessed.

By implication, yes. The NAEP legislation in effect just prior to the establishment of the Governing Board in 1988 included the requirement that the content to be assessed be defined. Specifically, the law required that "each learning area assessment shall have goal statements devised through a national consensus approach, providing for active participation of teachers, curriculum specialists, subject matter specialists, local school administrators, parents and members of the general public." (NAEP, 1988, p. 6) This process was used to develop the content-by-process matrix used for the assessments prior to the 1988 legislation, which are now largely referred to as the Long-Term Trend assessment (Mullins, 2017). The language related to assessment content in the current congressional authorization (P.L. 107-297, 2002) does not use the term "framework," but it has similar meaning.

What are the Legal Responsibilities of the Governing Board?

The responsibilities for the Governing Board as defined in the authorizing legislation (P.L. 107-297) are about more than developing assessment frameworks for NAEP. In Table 1 below, all of the requirements of the law are listed for clarity with the **distinctly framework-related ones shown in bold**. It should be noted that P.L. 107-279 is also about more than the Governing Board. It provides authorization for both the Governing Board (Section 302) and NAEP (Section 303). One requirement in Table 1 (No. 8) is from Section 303 and is included because it has implications for the policies and work for which the Governing Board is responsible. Also, references to Section 303 are found throughout Section 302 in acknowledgement of the necessity to coordinate all aspects of NAEP. While the requirements for the Governing Board in Table 1 are organized into an easier to read list than is typical presentations of laws, the correct legal citations are provided in brackets after each item.

Table 1 Legal Responsibilities of the Governing Board from P.L. 107-279 **(Emphasis added for distinctly framework-related responsibilities)**
1. There is established the National Assessment Governing Board which shall …" [Section 302(e)(1)] i. **formulate policy guidelines** for the National Assessment (carried out under section 303). [Section 302(e)(1)(A)] ii. **select the subject areas to be assessed** (consistent with section 303(b)); [Section 302(e)(1)(B)] iii. **develop appropriate student achievement levels** as provided in section 303(e); [Section 302(e)(1)(C)] iv. **develop assessment objectives** consistent with the requirements of this section and test specifications that produce an assessment that is valid and reliable, and are based on relevant widely accepted professional standards; [Section 302(e)(1)(C)] v. **develop a process for review of the assessment which includes the active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the public**; [Section 302(e)(1)(D)] vi. **design the methodology of the assessment to ensure that assessment items are valid and reliable**, in consultation with appropriate technical experts in measurement and assessment, content and subject matter, sampling, and other technical experts who engage in large scale surveys; [Section 302(e)(1)(E)] vii. consistent with section 303, **measure student academic achievement in grades 4, 8, and 12 in the authorized academic subjects**; [Section 302(e)(1)(F)] viii. develop guidelines for reporting and disseminating results; [Section 302(e)(1)(G)] ix. develop standards and procedures for regional and national comparisons; x. take appropriate actions needed to improve the form, content, use, and reporting of results of any assessment authorized by section 303 consistent with the provisions of this section and section 303; [Section 302(e)(1)(I)] and xi. plan and execute the initial public release of National Assessment of Educational Progress reports. [Section 302(e)(1)(J)]
2. The National Assessment of Educational Progress data shall not be released prior to the release of the reports described in subparagraph (J). [Section 302(e)(1)]
3. The Assessment Board may delegate any of the Assessment Board's procedural and administrative functions to its staff. [Section 302(e)(2)]
4. **The Assessment Board shall have final authority on the appropriateness of all assessment items**. [Section 302(e)(3)]
5. **The Assessment Board shall take steps to ensure that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias and are secular, neutral, and non-ideological**. [Section 302(e)(4)]
6. In carrying out the duties required by paragraph (1), **the Assessment Board may seek technical advice, as appropriate, from the Commissioner for Education Statistics and other experts**. [Section 302(e)(5)]
7. 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. [Section 302(e)(6)]
8. Such agreement (with the Secretary to participate in state assessments) shall contain **information sufficient to give States full information about the process for decision-making (which shall include the consensus process used), on objectives to be tested,** and the standards for random sampling, test administration, test security, data collection, validation, and reporting. [Section 303(b)(3)(B)(II)]

Have the legal requirements for frameworks changed over time?

The duties of the National Assessment Governing Board were initially authorized in the legislation establishing the Board in 1988 and have remained quite stable throughout periodic reauthorizations, the latest of which is P.L.107-279 (2002). This law provides authorization for both the Governing Board (Section 302) and NAEP (Section 303).

In each iteration of the law the subsections have been rearranged slightly and language was added, deleted or clarified. The requirements, however, have remained essentially the same. Two unique elements were added in 2002. The first was Section 302(e)(1)(D), [No. 1.v. in Table 1], which calls for an inclusive review process for the assessment that is now addressed both by a Governing Board policy (NAGB, 2002i)[4] and by the framework review/revision process involving panels of experts and the solicitation of public comments before each framework is adopted. The other addition was Section 302(e)(1)(F), [No. 1.vii. in Table 1], which provides a linkage to Section 303 – the NAEP section. Appendix B presents all of the legal requirements in a side-by-side arrangement. Each requirement is presented with the legal numbering used in each reauthorization and identifies changes that occurred in each revision.

III. Board Policy Work Impacting Assessment Frameworks

This section of the report takes a broad look at the policy work of the Governing Board and how these efforts have influenced the development of NAEP Assessment Frameworks and the Framework Development Policy.

Before the Governing Board Framework Policy

As noted previously, the 1990 NAEP Mathematics and Reading Frameworks were the first frameworks issued after the Board's establishment. These objectives initially were developed and published (1988 and 1989 respectively) under the NAEP National Assessment Planning Project. The project, just like NAEP in prior years, used the accepted professional practices for test development. However, this project was more political than previous NAEP assessments had been. That is, the opinions and endorsements of local and state education leaders became more important than ever before. As objectives-based assessments had grown in the states throughout the 1970's and 1980's, these leaders wanted to be sure that the NAEP assessments covered the content they considered important and that it was tested in ways they thought appropriate. Of course, NAEP had always considered the advice of the subject area experts, but the advent of state report cards heightened NAEP's importance to states and resulted in more scrutiny for the assessments. These leaders wanted to ensure that what was tested would be reflective of the essential content being taught in their schools.

[4] The Governing Board policy statement, *Review of the National Assessment of Educational Progress*, adopted August 3, 2002, included six guiding principles that describe expectations for the rigorous review of the National Assessment of Educational Progress and actions of the Governing Board.

Historical Processes Impacting Governing Board Policies

The Governing Board became an operational entity in October 1988 with six members from the existing Assessment Policy Committee and other members appointed to staggered terms by Secretary of Education William J. Bennett in September 1998. (Vinovskis 1998, p. 20) The first Board meeting occurred on November 18–19, 1988, just seven weeks after the law went into effect. Some of the first activities included hiring staff, establishing a way of work (adopting by-laws), and planning for the 1990 Reading and Mathematics Assessments. Two working groups (organizational and policy) were formed at the very first meeting of the National Assessment Governing Board, and work was begun to develop by-laws which were adopted a year later.

The early years of the Governing Board were spent addressing the responsibilities contained within the authorizing legislation, including plans for reporting, setting achievement levels, and preparing frameworks. Assessment frameworks were adopted in 1990, 1991, 1992, 1994, 1996, 2000, and 2001. The *Redesigning the National Assessment of Educational Progress Policy Statement* (NAGB, 1996) was adopted at a time when Congress had codified National Education Goals, and it was the expectation that the NAEP would be a primary means for monitoring progress in student achievement. The new National Education Goals called for more subjects to be assessed than in the past and, not surprisingly, assessment frameworks were addressed throughout the policy. Although the legislation has now been replaced by the *No Child Left Behind Act of 2002* (P.L. 107-097), some of the principles in that policy remain (e.g., inclusive process and stable frameworks).

The greatest impact on Governing Board policy development was the No Child Left Behind Act of 2002 (P.L. 107-097). That year was very busy and many policies were codified, including the *Framework Development* and *Item Development and Review* policies.[5] In his letter to Board members about the August 1-3, 2002 meeting, then Executive Director, Roy Truby, summarized these actions in the selected quotes which follow.

> Actually, the Governing Board's work on No Child Left Behind began more than a year ago at the Board's special meeting in Houston on June 28, 2001. It was then, … adopting the design changes that make it possible for 2003 to be the base year for the mandatory state NAEP. … At the March and May meetings, the Board adopted a new schedule of assessments, eight new policies, several changes in its by-laws, and one white paper to implement the law. At this meeting, three

[5] Governing Board policies codified after the passage of the *No Child Left Behind Act of 2002 included*: *NAEP and the No Child Left Behind Act* (NAGB 2001b), *Framework Development* (NAGB 2002a), *Item Development and Review* (NAGB 2002b), *Long-term Trend* (NAGB 2002c), *Plan for Study of NAEP Sampling* (NAGB 2002d), *Policies and Procedures for Complaints Related to the National Assessment of Educational Progress* (NAGB 2002e), *Prohibition on Using NAEP to Influence State and Local Standards, Tests, and Curricula* (NAGB 2002f), *Public Access to Test Questions, Item Release, and Confidentiality of Data for NAEP* (NAGB 2002g), *Resolution on Participation of the Commonwealth of Puerto Rico in NAEP* (NAGB 2002h), and *Review of the National Assessment of Educational Progress* (NAGB 2002i).

> more policies and a study plan have been prepared for Board action. (NAGB, 2002l)

A more complete history of the early days of the Governing Board can be found in the resource *Overseeing the Nation's Report Card* (Vinovskis, 1998).

Ongoing Governing Board Policy Work

Governing Board policies have operationalized the requirements in the law. They have, for example, determined how the work of setting achievement levels would be completed. Governing Board policy work is an ongoing activity and will require the attention of Board members and staff again and again.

Governing Board polices have been responsive to the law, but specific policies have not been required by the law. The need for a policy is solely determined by the Governing Board. As mentioned earlier, the *Redesigning the National Assessment of Educational Progress* policy included guidance related to framework development which is still being used today. The excerpts below are examples of Governing Board decisions to codify in policy topics that are not explicitly required in the law.

> Test frameworks and test specifications developed for NAEP generally shall remain stable for at least 10 years.
>
> In rare circumstances, such as where significant changes in curricula have occurred, the Governing Board may consider making changes to test frameworks and specifications before 10 years have elapsed.
>
> NAEP shall be designed so that others may access and use NAEP test frameworks, specifications, scoring guides, results, questions, achievement levels, and background data. (NAGB, 1996, pp. 14-16)

The Governing Board does continue to update policies. Recent examples, in addition to *Framework Development Policy*, are the *Reporting, Release, and Dissemination of NAEP Results Policy Statement* (NAGB, 2017a) and the policy on *Developing Student Achievement Levels for the National Assessment of Educational Progress* (NAGB, 2018c).[6]

Some policies originally established in 2002, such as the Framework Development Policy, have been updated but others have remained intact and are still relevant today. A primary example is the policy on the *Prohibition on Using NAEP to Influence State and Local Standards, Tests, and*

[6] Ongoing work on updating the *Item Development and Review Policy* (NAGB, 2002b) and the *NAEP Testing and Reporting on Students with Disabilities and English Language Learners Policy* (NAGB, 2010, 2014) has been severely impacted by the restrictions the COVID-19 Pandemic has imposed on the Governing Board and others across the country who would have participated.

Curricula (NAGB, 2002f). The law gave this admonition, but the Governing Board decided to codify its position in a policy.

Influence of Professional Standards

Implementing NAEP and Governing Board policy is not done in a vacuum. External influences such as changes in the content standards of professional organizations or the instructional practices for a content area are a consideration when developing or revising frameworks. For example, changes were made in the 1996 Mathematics Framework "which would better align the NAEP program in mathematics with the National Council of Teachers of Mathematics Standards (NCTM, 1989) and the Professional Standards for Teaching Mathematics (NCTM, 1991)." (NAGB, 1992, p. 2) Another example was the nationwide emphasis on the preparedness of high school graduates for the workplace and college. A review of the mathematics and reading assessment frameworks was conducted and changes were made. (Achieve, 2005; Achieve, 2006)

There are also professional standards in the field of tests and measurements, known as psychometrics. As the Governing Board has developed policies, the staff and contractors have worked to adhere as closely as possible to these standards and also to the statistical standards of the National Center for Education Statistics. Both editions of the Framework Development Policy make reference to the following standards. The 2018 edition of the policy states it this way. (NAGB, 2018b)

> This Policy complies with the National Assessment of Educational Progress Authorization Act of 2002 (P.L. 107-279) and the documents listed below which express widely accepted technical and professional standards for test development. These standards reflect the agreement of recognized experts in the field, as well as the policy positions of major professional and technical associations concerned with educational testing.
>
> *The Standards for Educational and Psychological Testing*. (2014). Washington, DC: American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.
>
> *Code of Fair Testing Practices in Education*. (2004). Washington, DC: Joint Committee on Testing Practices.
>
> *Center for Education Statistics (NCES) Statistical Standards*. (2012).

These standards emphasize features of tests including, for example, the content to be assessed and the statistical information that should be provided about test items and tests as a whole. If these standards are updated, the Board must work to address any new components that are applicable to NAEP and update the Governing Board policies, practices, and procedures, as may be needed. Contractors are expected to implement framework development projects in a manner that honors and is congruent with these standards. The requirements document for

the most recent frameworks procurement describes the procedures expected of contractors so that an assessment consistent with the standards will be implemented. (NAGB 2018a)

One challenge should be noted. The documents cited above focus primarily on the assessment and reporting of individual student scores. NAEP does test individual students but does not report individual scores. Thus, the professionals working in these areas must interpret how these standards are intended to apply to the unique situation of NAEP. While these standards are updated from time to time, it is infrequent. The most recent editions emphasize collecting many types of validity evidence in order that the validity claims of an assessment can be supported. Validity has always been important to NAEP and the Governing Board, and to the organizations which have evaluated NAEP. (National Research Council, 1999; Buckendahl, et.al., 2009; National Academies of Sciences, Engineering, and Medicine, 2017) Therefore, collecting validity evidence for NAEP and implementing other applicable portions of the standards will continue to be an important consideration for the Governing Board. In this regard, the Board examines the overlap between the NAEP framework and the standards used by other organizations and states. Recently, comprehensive reviews of state standards were conducted for mathematics and science. (AIR, 2018a, 2018b, 2018c, 2018d; HumRRO 2021)

IV. Board Policy for Framework Development

This section of the report focuses on the Governing Board Framework Development Policy, its origins, components, and changes over time. In addition, a list of Board decision points for framework development are presented.

2002 Framework Development Policy

The first Framework Development Policy was adopted on May 18, 2002 (NAGB, 2002a). As described earlier, the framework development activities conducted from 1988 to 2002 utilized processes similar to those codified in 2002. In particular, an iterative process was followed that used committees of content specialists from the field, a consensus process, opinions solicited from stakeholders, and the involvement of the Governing Board. The intent of the Assessment Development Committee (ADC) to incorporate similar guidance into the policy is manifest in their March 1, 2002, meeting minutes. (NAGB, 2002i)

> … the Executive Committee delegated this issue to the ADC since it involved the area of framework development and item review. ADC members discussed the current Board practice of "casting a wide net" to have broad representation on the framework development panels. The new policy language should make this explicit, perhaps by setting targets for representation of various NAEP constituencies. Strategies for involvement and feedback from the general public should also be stipulated. A draft policy will be prepared for discussion at the May Board meeting. (NAGB, 2002j)

At the May 2002 meeting, the Governing Board reviewed the policy ADC recommended for adoption. The ADC minutes of that meeting contain the following statements.

> This policy was reviewed and discussed in detail at the ADC's April 29 meeting in Detroit, Michigan. Committee members had no further changes to the draft policy. Action Item: The Assessment Development Committee recommends Board approval of the Policy on Framework Development. (NAGB, 2002l)

After receiving the ADC report and recommendation, the first Framework Development Policy was adopted. (NAGB, 2002a) The purpose of establishing this policy was to incorporate the requirements of the authorizing legislation and professional best practices into an official policy that provided explicit guidance for Governing Board staff and contractors to follow in framework development projects. The original 2002 policy was organized around seven principles with additional guidance about how to implement each of the principles. Simply stated, the policy provided for the following.

Principle 1 – the definition of a framework and what is to be included

Principle 2 – the process and participants for developing the frameworks

Principle 3 – the inclusion in the review process of current theory and practice standards within the discipline as defined by a variety of organizations

Principle 4 – the role of the Governing Board in approving the framework and the role of its designees including committees, staff, and contractors that might be hired by the Governing Board, and the required documents to be presented to the Board for approval

Principle 5 – the inclusion of preliminary achievement level descriptions and intended uses of them

Principle 6 – specific instructions, to be used by others, for the design of the test and constructing items

Principle 7 – the expectation that frameworks would remain stable for at least 10 years

2018 Framework Development Policy

In 2018, the Governing Board made a revision to the 16-year-old Framework Development Policy. (NAGB, 2018b) In addition to some minor reorganization and rewording, primary distinctions between the 2002 and 2018 editions included four changes that will be discussed in this section: (1) updating frameworks, (2) reviewing frameworks, (3) participants/stakeholders, and (4) framework panels/committees. Additionally, the current policy maintains a focus on the overarching principles to be followed, with the details and procedures moved to procedural documents and requirements for contractors. (NAGB, 2018a)

This section first describes the general contents of the 2018 policy and subsequently provides more detail about the four changes mentioned above. The two versions have similar content, although they are arranged somewhat differently. Appendix C contains a more detailed comparison of the policy principles for both versions in a side-by-side display. Although Appendix C does not capture all of the edits which occurred to remove redundancy and procedures, it does provide some examples of the specific wording changes.

The 2018 policy was organized around six principles, each containing additional guidance about how to implement the principle. Simply stated, the policy provides for the following.

Principle 1 – Elements of Frameworks: the scope of the domain to be measured, delineating the knowledge and skills to be tested at each grade, the format of the NAEP assessment, and the achievement levels. (Note: Combines 2002 Principles 1 and 5.)

Principle 2 – Development and Update Process: develop and update frameworks through a comprehensive, inclusive, and deliberative process that involves active participation of stakeholders. (Note: Updating frameworks was added to this section.)

Principle 3 – Framework Review: determine whether an update is needed to continue valid and reliable measurement of the content and cognitive processes reflected in evolving expectations of students and anticipates a framework review at least once every 10 years. (Note: This section was added to describe the process for determining if a framework update is needed and to address timing included in 2002 Principle 7.)

Principle 4 – Resources for the Process: take into account state and local curricula and assessments, widely accepted professional standards, exemplary research, international standards and assessments, and other pertinent factors and information.

Principle 5 – Elements of Specifications: shall be developed for use by NCES as the blueprint for constructing the NAEP assessment and items.

Principle 6 – Role of the Governing Board: shall monitor all framework development and updates. The result of this process shall be recommendations for Governing Board action in the form of three key documents: the framework; assessment and item specifications; and contextual variables that relate to the subject being assessed.

Updating Frameworks. The original Framework Development Policy in 2002 was stated in terms of developing new frameworks because this had been the primary focus of the work at the time the policy was adopted. Only Principle 7 referred to revising frameworks, but provided little guidance about the process. Therefore, the 2018 revision of the original policy was undertaken to include provisions for updating frameworks when a complete revision might

not be necessary. References to updating frameworks were added throughout the policy and guidance about the update process was included in Principle 2.d.

> The scope and size of a framework development project shall determine the size of framework panels and the number of panel meetings needed. A framework update project may require smaller panels and fewer meetings if a smaller scope is anticipated for recommended revisions. Each project shall begin with a review of major issues in the content area. For a framework update, the project shall also begin with an extensive review of the current framework, and the Visioning Panel shall discuss the potential risk of changing frameworks to trends and assessment of educational progress. (NAGB, 2018b, p. 6)

An important consideration for making decisions to update a framework is the potential impact on NAEP reporting. This concern was addressed under Principle 6.d. "In initiating a framework update, the Governing Board shall balance needs for stable reporting of student achievement trends. Regarding when and how an adopted framework update will be implemented, the Board may consider the NAEP Assessment Schedule, cost and technical issues, and research and innovations to support possibilities for continuous trend reporting." (NAGB, 2018b, p. 9)

Reviewing Frameworks. In the 2018 Framework Development Policy, a process was included for reviewing frameworks to determine if/when an update was needed. Principle 7 of the 2002 policy emphasized the importance of holding a framework stable for 10 years. The 2018 new Principle 3 calls for reviewing frameworks at least once every 10 years. Further, this new principle describes the review as considering the current relevance of the assessments and frameworks, input from experts, and the risk of changing the reporting of trends. The policy makes clear the decision to update involves the full Board's recommendation and describes the process for conducting an approved update.

Principle 3 also explains that ADC, within the 10-year period, may observe major changes in the states' or nation's education system related to NAEP frameworks and when/if these changing conditions warrant recommending an update to the full Board. The Board's decision may involve convening a Visioning Panel to examine the issues including commissioning special research and analysis to inform the updates under consideration. Based on these findings, a determination will be made about next steps and the processes to be implemented as described in the policy.

Participants/Stakeholders in Framework Panels. The 2018 policy identifies the various stakeholders in a comprehensive list (page 2) that applies to all aspects of the framework development or update processes. In the 2002 policy, stakeholders were identified under various principles and consistent terms were not always used. The 2018 policy, also provides more specificity about the participants in the framework development panels. While both policies call for the use of content experts, curriculum specialists, state and local educators, and policy makers, the 2018 policy is more specific about involving members with classroom teaching experience. The 2018 policy specifies that at least 20% of the members have

classroom teaching experience, perhaps in recognition that it may be difficult for current classroom teachers to make the time commitments required for these projects, even though funds for substitute teachers are included. For example, a recent framework project required approximately 15 days of meetings. The bottom line as described in the contract requirements document is that anyone chosen to serve on these panels "must be well qualified by content knowledge and familiarity with the knowledge, skills, and abilities in the respective subject, while addressing all grade levels designated for the assessment." (NAGB, 2018a, p. 16)

Additionally, the 2018 policy identified an upper limit for the number of participants in panels. Although the 2018 policy does not provide a rationale for these limits, perhaps this change was to facilitate the consensus process, as well as shorten timelines and reduce expenses. The number of panel members working on past projects has sometimes been much larger than 30. For example, the project for the 2009 NAEP Science Framework development used a total of 57 panelists, with no duplication across committees. A challenge with using only 30 panel members will be to attain the desired diversity for the framework panels as described on page 5 of the policy (NAGB, 2018b). Balancing these competing priorities will be an ongoing consideration. Fortunately, the 2018 policy recognizes that it may be necessary to add additional members. This option will be most needed for projects that are large in scope, that is, all three grade levels and multiple areas of expertise required.

It should be noted that the participants in framework development panels are identified by the contractor hired to conduct the assessment development activities. This is not a nominations process. Governing Board staff (sometimes Governing Board members) review the proposals and monitor the implementation of contract activities. For example, if the diversity or classroom experience goals indicated in the policy are not present in the names submitted as panelists, staff would ask the contractor to augment the panel to account for identified deficiencies.

Table 2, which is found at the end of the next section, includes a summary of the stakeholders discussed in this section and their expected panel assignments.

Framework Committee/Panel Functions. The 2002 and the 2018 policies are both nominally and substantively different: nominally in terms of the panel names and substantively in their composition. Both policies utilize two framework development groups and they have separate functions – the first function is to develop the high-level guidance for the work and the second function is to develop drafts of the documents that are consistent with the guidance. The more substantive difference is their composition and division of labor. The 2002 policy provides for separate groups of individuals and the 2018 policy provides for overlapping participants in the visioning and development activities. Although the policy does not specify the rationale for the overlap, it is likely the development panel will more fully understand the vision and guidelines for completing the work without having to be informed about it separately.

A third group of panelists is the technical advisors, primarily testing specialists. The 2018 policy describes their involvement as a resource to the framework development work rather than as a

committee. This approach permits different experts to be involved on different topics when their expertise is needed. For example, expertise about assessing certain types of content or expertise about the impact of changes on maintaining trends. The framework panels would be able to get expert advice as needed during their deliberations rather than waiting for a meeting of the technical advisors to be scheduled. The work of the technical advisors is expected to be conducted by representatives who participate in framework development meetings and as a group in separate meetings for more in-depth technical discussions.

Table 2 below provides a comparison of the functional working groups and the participants in each which were discussed in the previous sections.

Table 2
Framework Development Groups Comparison

2002 Policy (NAGB 2002a)	2018 Policy (NAGB 2018b)
Policy Oversight/Steering Committee • Represents key policy groups, etc. • At least 30% users and consumers • Formulates guidelines for the process consistent with law and NAGB charge • Monitors progress of project • Reviews final product before Governing Board	**Framework Visioning Panel** • Represents all stakeholders, including policy makers and users/consumers • At least 20% have classroom teaching experience • Formulates initial guidance for framework development • Includes up to 30 members (including up to 15 on Development Panel) • Additional members as needed
Planning Committee • Content experts & educators, etc. • Consider NAGB Charge and project guidelines • Develop deliverables • No overlap with Steering • Classroom teachers "well represented"	**Framework Development Panel** • Subset of Visioning Panel • Proportionally higher content experts & educators than the Visioning Panel • Detailed deliberations to resolve issues & recommend framework • Up to 15 members • Additional members as needed
Committee of Technical Experts (TAC) • Primarily testing experts • Involved where appropriate • Respond to technical issues raised by the committees • Review documents, esp. specifications • Provide guidance to project staff	**Technical Experts (TAC)** • Primarily testing experts • A resource to framework panels • Respond to technical issues raised during deliberations and meet separately, as needed • Review documents, esp. specifications

Natural Tension Points

The Framework Development Policy recognizes several natural tensions that exist in the education community at large. Education disciplines and the professionals who work within them are not unidimensional. Professionals naturally have different viewpoints about what is most important, what is most important to assess, and how that content should be assessed and reported. The policy provides the following guidance about the consensus process for developing or updating an assessment framework as broadly inclusive as possible.

> In balancing the relative importance of various sources of information, framework panels shall consider direction from the Governing Board, the role and purpose of NAEP in informing the public about student achievement, the legislative parameters for NAEP, constraints of a large-scale assessment, technical assessment standards, issues of burden and cost-effectiveness in designing the assessment, and other factors unique to the content area. (NAGB, 2018b, p. 8)

Additionally, there are frequently concerns about the scope of the content or range of content difficulty included in a framework. The Framework Development Policy recognizes this as natural tension point and provides the following guidance about addressing this concern and resolving it through the panel consensus process.

> The NAEP framework development and update processes shall be informed by a broad, balanced, and inclusive set of factors. The framework shall reflect current curricula and instruction, research regarding cognitive development and instruction, and the nation's future needs and desirable levels of achievement. This delicate balance between "what is" and "what should be" is at the core of the NAEP framework development process. (NAGB, 2018b, p. 7)

These are not all of the possible tension points that can arise in a broad-based committee process where varying opinions naturally exist. However, they do illustrate the Board's acknowledgment of them and guidance about resolving issues when they arise.

Resolving Points of Disagreement

Clearly, the Board acknowledges that different people and groups have different opinions about even the simplest constructs. In every framework adoption process, there is always some disagreement about the decisions represented in framework documents. The Framework Development Policy anticipates that there will be differences of opinion and provides guidance in this regard.

> Panels shall consider all viewpoints and debate all pertinent issues in formulating the content and design of a NAEP assessment, including findings from research.

> Reference materials shall represent multiple views. For each project, protocols shall be established to support panel deliberations and to develop a unified proposal for the content and design of the assessment. Written summaries of all hearings, forums, surveys, and panel meetings shall be made available in a timely manner to inform deliberations. (NAGB, 2018b, p. 6)

This is not a new challenge. Resolving these differences is what was envisioned by use of the term "consensus process" in the authorizing legislation. As mentioned in an earlier section of this report, the very first Reading Framework contains this statement.

> While objectives resulting from such a consensus process reflect neither a narrowly-defined theoretical framework nor every view of every participant, they do represent the thinking of a broad cross section of individuals who are expert in the areas of literacy research and reading instruction and who are deeply committed to the improvement of reading in our schools. (NAGB, 1990, p. 8)

Another example is the statement made by Charles Smith, then Executive Director, at the August 2004 Board meeting about the adoption of the 2009 Reading Framework which was two years in the making.

> Thousands of hours of effort have been devoted to the initiative, and the result awaiting your decision is, I understand, the most scrutinized framework ever to come before this Board. (NAGB, 2004e)

As the Governing Board has become more experienced in the process of identifying the content to be assessed, the framework documents themselves have become more thorough and more thoroughly and openly discussed. The Governing Board has expanded the involvement of experts in the field, utilized the research base within each discipline, and provided more opportunities for public comment. These activities are discussed in the next section of this report.

V. Framework Development and Implementation Activities

The legislation and Framework Development Policy have not changed substantially since enacted, but the activities to implement a new framework or update an existing one are much more extensive today than they were in the early 1990's. Some of the important changes are highlighted in this section.

Developing and Updating Assessment Frameworks

The development of a framework for a new assessment or updating one is guided by the schedule of NAEP assessments adopted by the Governing Board. (NAGB, 2018b) The

assessment schedule is a forward-looking document and identifies when changes in a framework might be expected. When development of a new framework or a framework update is initiated, several concerns must be balanced. For example, the need for stable reporting of student achievement trends, cost, specific changes in the discipline, relevant research, and innovations or new initiatives in impacting the field. These concerns are mostly objective considerations, but there are also more subjective elements. For example, when the subject area includes competing ideologies for which there is no obvious consensus, it can lengthen the timeframe for completing the framework. Making a decision to develop or update a framework is a complex process and involves many decision points as discussed in the following section.

Framework Decision Points

The framework policy broadly describes the process for developing a new framework and updating an existing one. It does not prescribe an order of events, although one may be logically inferred from the policy. Throughout the process of framework development, there are a number of important interactions between the Governing Board and its committees, subject area experts, stakeholders, the general public, and the panels convened to make recommendations to the Board.

The Governing Board by-laws assign responsibility for implementing the processes involved in framework development to the Assessment Development Committee (ADC). Their duties in this area include: developing and implementing a broadly inclusive process, developing content objectives, ensuring the active participation various stakeholders, developing assessment specifications, and providing for the review of test frameworks and specifications by other groups. (NAGB, 2010b, page 7) Additionally, the by-laws assign to ADC the responsibility of reviewing subject-specific background questions and all cognitive test items.

Consistent with the by-laws, Principle 6 of the 2018 policy describes the role of the Governing Board and ADC for framework development. (NAGB, 2018b, page 9) ADC's role is to monitor all the activities leading up to a framework development or update project and the ongoing project work. The Board's role is to approve and adopt the charge to the Visioning Panel and final framework documents prior to their handoff to NCES for developing the test questions. Although the Assessment Development Committee has the primary role for oversight of framework development/updating processes, other committees of the Board and NCES are involved as needed. Typically, COSDAM is involved in technical issues (scoring, scaling, trend reporting, etc.), R&D is involved in discussions about reporting and contextual data collection, and NCES is involved in issues related to item development, test construction, test scoring, data analysis, and reporting.

The discussion below provides a brief summary of important decision points and offers fundamental questions to be answered during the process of developing or updating a framework. It does not include every possible question or interaction between the Board, its committees, and other organizations. Appendix D supplements the information provided

below with a little more detail about the range of actions and the involvement of the Board, the Assessment Development Committee, contractors, and external reviewers.

1. **Should a framework revision or update be considered?** At least once every 10 years the Assessment Development Committee determines the timing for review of frameworks based on two key variables – the NAEP Assessment Schedule and lead time needed to implement a new/revised framework, including developing and field-testing new items for the assessment. The committee considers the relevance of assessments and their underlying frameworks, and any changes occurring in the field in making this decision. In their deliberations, the Assessment Development Committee may solicit input from experts, hear testimony or review white papers, discuss and determine what action should be recommended to the full Governing Board. Recently, comprehensive reviews of state standards were conducted for mathematics and science to document the overlap between the NAEP frameworks and the array of State standards before deciding to pursue a framework update. (AIR, 2018a, 2018b, 2018c. 2018d; HumRRO 2021)
2. **Is a new framework or update needed?** The Board receives a report from the Assessment Development Committee about their discussion and recommendations about the framework. Depending on the issues and interest, the Board may also hear presentations from various experts. If the Board agrees with the Assessment Development Committee recommendation, they will review, revise (if needed), and adopt the charge to the Visioning Panel. Many other actions will follow including contracts, working panels, and revised framework documents. See Appendix D for additional detail on these activities.
3. **Is the draft framework ready to be evaluated by external reviewers?** As the work to develop the framework proceeds, Governing Board staff carefully monitor the entire process. They have weekly conference calls with the project team and attend all the meetings of the Visioning and Development Panels. Others also attend the panel meetings, including the project technical advisors and representatives from NCES. This involvement throughout the project identifies and resolves potential issues. The Assessment Development Committee receives regular reports from the Framework Development Project staff and Governing Board staff, who in turn provide updates to and seek input from other Committees of the Governing Board on issues related to their areas of expertise and responsibility. Governing Board staff, in consultation with the Assessment Development Committee, determine when the contractor can begin the process of conducting external reviews. Agreements with the contractor describe how feedback will be solicited, reviewed, and incorporated.
4. **What feedback should be incorporated in the Framework?** The Framework Development Panel must consider all viewpoints, debate all pertinent issues about the content, including findings from research, and make revisions to the framework accordingly. This will likely be an iterative process, that is, reviewing and revising framework documents may occur more than once. After feedback is incorporated, the final draft is shared with staff and the Assessment Development Committee who review and recommend revisions or approval by the full Board.

5. **Should the framework be adopted and implemented?** In making a final decision, the Board should consider the process used to develop the framework, the role and purpose of NAEP to inform the public about student achievement, the legislative parameters for NAEP, constraints of a large-scale assessment, technical assessment issues (for example, the continuation of trend lines), issues of burden and cost-effectiveness in designing and implementing the assessment, and other factors unique to the specific content area. After the framework is approved, the next logical steps will be the development of item specifications and contextual variables for the assessment. Although it is likely the Panels have been considering these elements throughout their deliberations, they will formalize a document containing the prescribed information and submit it to the Board for review and approval through the Assessment Development Committee. Once approved, NCES and their contractors will begin item development and other planning for the assessment.

Appendix D supplements the information provided above with a little more detail about the range of actions and the involvement of the Board, ADC, contractors, and external reviewers. It highlights the major questions/decisions and other subordinate ones needed for framework development, approval, and adoption by the Board. Many smaller decisions and steps are behind these major decision points, but cannot be captured in this simplistic presentation. While the decision points are presented in an orderly manner, they may not always be implemented in the chronology implied by this list.

Need for Subject Area Updates

The 2018 Framework Development Policy added an entire section on how framework reviews would be conducted. For example, "the ADC shall solicit input from experts to determine if changes are warranted, making clear the potential risk of changing frameworks to trends and assessment of educational progress." (NAGB, 2018b, page 6) In making a decision about updating a framework, the Board needs to have explicated how extensive the revisions to a framework are likely to be, e.g., if substantive change would be required in the content being reported. For example, a major change would be changing the content areas and subscores reported. A more minor update could keep the test design and reporting intact, but recommend changes in how the content is assessed or which elements of the content are no longer relevant. Obtaining clarity about the need for an update in a subject area could involve the solicitation of white papers from subject matter experts about how the subject area should be assessed and important elements that should be considered. Another alternative could involve a panel discussion at an Assessment Development Committee or a full Board meeting. In either case, it will be the Board's responsibility to determine if a revision or update is needed.

Framework Panelists

The Board has always valued the opinions of and made every attempt to include classroom teachers, curriculum specialists, school administrators, policy specialists, subject-matter experts, and representatives of the general public in framework development projects.

However, balancing the membership of panels is not easy. The current Framework Development Policy provides the following guidance.

> In accordance with the NAEP statute, framework development and update processes shall be fair and open through active participation of stakeholders representing all major constituents in the various NAEP audiences, as listed in the introduction above.
> Framework panels shall reflect diversity in terms of gender, race/ethnicity, region of the country, and viewpoints regarding the content of the assessment under development. (NAGB, 2002a, pg. 5)

The role of the Governing Board, in particular the staff, and the Assessment Development Committee, is to review the panelists recommended by the contractor and ensure they meet the rigorous requirements of the contract. "All panelists must be well qualified by content knowledge and familiarity with the knowledge, skills, and abilities in the respective subject, while addressing all grade levels designated for the assessment." (NAGB, 2018a, p. 16) If there are concerns about panelists individually or collectively, it is incumbent upon the Governing Board to communicate these concerns and ensure they are addressed promptly.

The Framework Development Policy adopted in 2018 made some changes to the composition of the panels. Please refer to that earlier section for those details.

Public Comment Opportunities

It has always been the practice of the Board to seek public comment on the framework to be adopted. Sometimes, this included only advertising a comment opportunity in the Federal Register which may have limited the number of comments received. Since the early 2000's, the Board has expended much more effort in seeking feedback. Examples include public forums, meetings with state leaders in the content area and assessment directors, and working collaboratively with policy advisory groups and professional associations. The current policy guidelines emphasize the importance of a broad reach in obtaining public comment.

> Public comment shall be sought from various segments of the population to reflect many different views, as well as those employed in the specific content area under consideration. (NAGB, 2002a, pg. 5)

People who comment on a framework usually represent a constituency and have a particular viewpoint to be expressed. Their opinions may be minute or major and may be raised quietly or loudly. No matter, their opinions are important and hearing them is important. This does not mean the Governing Board is compelled to implement all recommendations made during the public comment period.

Constraints – Cost, Contracting, and Timelines

In addition to the decision about developing or updating a framework, the Governing Board must also contend with matters of budget, contracting, and timelines. These concerns are interrelated and difficult to parse.

Cost Factors. The Governing Board budget is constrained by the appropriation of funds from Congress. The cost of a framework development project depends on a number of factors including the complexity of the requirements, the competitiveness of the marketplace, the timeframe for completing the project, the extensiveness of revisions requested, and the unexpected. As might seem obvious, the more complex the project and the longer it takes to complete, the more expensive it will be. Some of these factors are predictable, but others, like the COVID-19 pandemic, are more are difficult to anticipate. In general, the Governing Board budget is sufficient to cover the cost of developing new or updating existing frameworks when done one at a time. Circumstances requiring multiple contracts in the same year may entail extensive advance planning to accommodate.

Framework Contracts. Contracts with organizations experienced in developing educational assessments have been used by the Board since it was established in 1988. The very first frameworks were supported by contracts with the Council of Chief State School Officers (CCSSO) that established the National Assessment Planning Project. Over the history of framework development, contracts have been awarded to the American Institutes for Research; American College Testing; the College Board; the Council of Chief State School Officers; the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA; and WestEd, and others. (Jago, 2009)

In recent years, the number of contractors bidding on NAEP Assessment Framework Development contracts has dwindled. The failure to have multiple bidders is a disadvantage because choice in vendors is desirable, as is competitive bidding. The root cause of the reduction in bidders is unknown, but reasons can be assumed to include the uniqueness of the project, lack of prior experience, changing or realigned corporate capabilities, availability, conflict of interest, potential for controversy, lack of interest, or other factors.

Contracting Procedural Requirements. The sophistication of the framework development procedures and contracting requirements has grown over time. The Framework Development Policy implies a number of processes that should be completed by those developing frameworks, but the contract requirements are much more detailed. For example, the policy is contained in nine pages, but the current Governing Board procedural requirements for contractors is 35 pages long. These requirements were recently Attachment A to the Governing Board procurement *Update of National Assessment of Educational Progress (NAEP) Frameworks for Mathematics, Reading, and Other Subjects*. (NAGB, 2018a)

The length is necessary because of the number of detailed requirements contained therein. The current work calls for regular monitoring of the project by Governing Board staff, and

regular reporting to the Assessment Development Committee throughout the scope of the contract. Attention is also given to the identification of panel members and the processes being implemented. A process report is required which summarizes all procedures implemented and issues encountered. This detailed information is used to support the validity of the recommended framework, specifications, and contextual variables. The Table of Contents from the most recent Statement of Work is found in Appendix E and shows the extensiveness of the requirements covered.

Timelines. This discussion about timelines will be considered from two perspectives: the time required to develop and adopt a new framework, and the lead time to implement changes to the assessment. These are related in that the latter cannot be accomplished without the former.

The lead time for changes to the assessment will be considered first because it has a fixed end point because of the NAEP assessment schedule. According to information NCES has communicated at Board meetings, the timelines for creating new assessment items and including them in a NAEP assessment can take from five to six years, whether the assessment framework is new or is being updated and applies equally to developing a new digital-based assessment or digital items for an existing assessment. This timeline is long because items must be developed and reviewed, tried out with small groups of students, analyzed, added to existing assessments, and then administered in an actual NAEP assessment. Because NAEP is not administered every year this timeline is longer than is typical for most assessment programs.

In understanding this timeline, it might be helpful to think about developing assessment items in three phases.

- The first phase is to develop questions for cognitive skills to be assessed, including reviews by experts in the field and conducting cognitive labs to ensure the questions are assessing the cognitive skills intended by the framework. Sometimes, several rounds of review and revision are needed to develop questions that meet the NAEP framework and review criteria. These questions also must be formatted for the platform on which they will be presented and reviewed in that same manner.
- The second phase involves collecting data from students which is called pilot testing. This is usually done during a regular NAEP testing window. Questions for this phase must be formatted and presented as they ultimately will appear on NAEP. Sufficient quality control steps must be performed to ensure data capture and scoring are accurate. Additionally, data must be collected from a significant number of students so that results can be correctly interpreted and used to develop future forms of NAEP. Another round of reviews occurs after these data are collected which includes examining item and test statistics, including item bias. If questions are rejected at this point, they may be revised and recycled through the first two phases.
- The third phase involves administering forms (blocks) in the actual NAEP assessment, administering them to students, scoring questions, and summarizing the data to be reported.

The schedule may also depend on when the Board authorized the work to begin as well as the level of innovation represented in the items identified in the framework. After the Governing Board approves the assessment framework, item specifications, and contextual variables, work can begin. After item writing is completed and items are reviewed by standing committees of content experts and the Governing Board, the approved items can be field tested (item tryouts) with the target group of students. Field testing will be done during the regular NAEP assessment window with a special sample of students. Those items which survive statistical standards and another round of reviews are assembled into forms and reviewed by NCES and the Governing Board. Because the field testing is done in one calendar year and the actual test administration is done in another, the minimum amount of time needed is two years. However, if new item types or constructs are contained in the framework, or if an innovative delivery of item content must be explored, more time will be required to try out items and analyze them before they are deemed valid for their intended purpose. It is not the purpose of this paper to discuss cognitive labs or other methodologies useful in determining item validity. It is enough to say this takes much longer.

The most obvious statement to be made about developing frameworks is that developing a new framework should take longer than updating an existing framework; however, that statement is very misleading. The more agreement there is in a subject area is probably a better factor for predicting how much time will be involved in developing a new framework or updating an existing one. As the Framework Development Policy prescribes, the Governing Board is seeking a consensus project; therefore, the longer it takes to reach consensus the longer the framework project will take. In thinking about the timeline for a framework project, one cannot think only about the framework panels who make content recommendations to the Board. One also must consider the time required to hire contractors on the front end of the work, as well as the public comment period and Governing Board deliberations/actions on the back end. In the best-case scenario where there is a great deal of consensus about the content to be assessed and when the public commentary is also agreeable, a period of one to two years can be expected for developing a charge, issuing a procurement, hiring a contractor, convening panels, etc. In the worst-case scenario where there is contentious debate, much more time is required. Finally, if the Board cannot support the recommended framework and reach a compromise that the Visioning and Development panels can support, then the entire process must begin again.

VI. Issues for the Future

In recent years the Governing Board has been having strategic discussions and reflecting on the data NAEP has been reporting over the last 40+ years. These discussions were designed to focus the Board's work on the strategic priority of providing NAEP information in the most innovative and effective ways. The Governing Board Strategic Vision for 2020 was adopted in November 2016 and the Strategic Vision for 2025 was adopted in September 2020 (NAGB, 2020b). Both of these efforts have included a vision for assessment frameworks. In both vision

statements, the reference to frameworks is found in the goal area "to innovate." Both versions are shown below with emphasis added.

2020 Strategic Vision
The National Assessment Governing Board will revise the design, form, and content of The Nation's Report Card using advances in technology to keep NAEP at the forefront of measuring and reporting student achievement. The Governing Board will 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.
2025 Strategic Vision
The National Assessment Governing Board will ensure The Nation's Report Card remains at the forefront of assessment design and technology by refining design, content, and reporting, increasing relevancy for NAEP users and inspiring action to improve achievement for all. The Governing Board will optimize the utility, relevance, and timing of NAEP subject-area frameworks and assessment updates to measure expectations valued by the public.

As the Board continues implementing their Strategic Vision for 2025, they will establish priorities for the ongoing assessment framework activities. Consequently, discussing the issues about future framework development seems appropriate in this paper.

Framework Responsiveness

For the development of the Board's 2020 Strategic Vision described above, work groups were formed to consider avenues for advancing NAEP. These working groups and committees explored new approaches that could be utilized. One of the discussions focused on how the NAEP frameworks could become more responsive to small changes in the discipline area. The aim was to make adjustments in a manner that could reduce the timeframe typically required to change a NAEP framework and assessment.

At their joint "strategic vision" planning meeting in November 2016, the Assessment Development Committee (ADC) and the Committee on Standards, Design, and Methodology (COSDAM) discussed the concept of making the frameworks more responsive to the current curriculum standards being implemented on a broad scale (e.g., the Common Core State Standards). Other topics discussed included maintaining trends, valid alignment with student learning activities (e.g., writing using word processing), lead time for changes, the extent of NAEP's alignment (or lack thereof) with state and other content standards, changes in the field that might not be detected by the static nature of NAEP, communicating incremental changes to the public, not creating moving targets for school systems, and the concept of dynamic frameworks. (NAGB, 2016) (NAGB, 2017, p. 36)

At the joint meeting of these two committees in March 2017, there was a more in-depth discussion of the "dynamic framework" concept. The Governing Board committees agreed that the term "dynamic frameworks" was not the best way to characterize this effort because it implied that the frameworks would constantly be in flux, and such fluidity or the perception of it could have unintended consequences as well as miscommunicate the nature of the updates which might occur. There also was agreement that more discussion and study about this topic was important with the goal of learning how frameworks could become more responsive without affecting NAEP's trend reporting. (Haertel, et.al., 2012, pp. 3, 16-17) (NAGB, 2017, pp. 28-29)

The concept of "dynamic frameworks" as presented in the *Future of NAEP Panel White Paper*, is intriguing. The paper suggests these considerations.

> Dynamic frameworks would balance dual priorities of trend integrity and trend relevance. … it would be important to establish and to enforce clear policies concerning the reporting of significant changes in assessment frameworks, so as to alert stakeholders when constructs change and to reinforce the crucially important message that not all tests with the same broad content label are measuring the same thing. (Haertel, et.al., 2012, p. 17)

This discussion is ongoing.

Standing Subject-Matter Committees

Another idea for identifying changes needed in a framework is to make use of NAEP standing subject-matter committees. NCES contractors establish standing committees of content experts, state and local education agency representatives, teachers, parents, and representatives of professional associations to review the items developed for NAEP. Each standing committee considers: the appropriateness of the items for the particular grade; the representative nature of the item set; the match of the items with the framework and test specifications; and the quality of items and scoring rubrics. (NCES, 2020b)

The *Future of NAEP Panel White Paper* makes the case for using such committees as follows.

> Under our proposal, standing committees would review field test data, for example, and be aware when "after-the-fact" distortions of the intended domain occur because more ambitious item types fail to meet statistical criteria. Standing committees could also update assessment frameworks incrementally, at the same time assuring that the constructs underlying NAEP reporting scales did not drift to the point where new trend lines were indicated. In particular, assessment frameworks would be updated to accommodate changing learning environments. Inquiries with dynamic knowledge representations and simulations in science would be one example. (Haertel, et.al., 2012, pp. 17, 44)

The NAEP contractors already use standing subject-matter committees, particularly for item reviews. However, they are not charged with the explicit functions described by Haertel, et al. It is customary for Governing Board staff to attend the debriefing sessions of these committees, so some consideration could be given to seeking input as suggested.

Digital-Based Assessment Frameworks and Policy

NAEP transitioned to digital based assessments in 2017. Updating frameworks in this context should provide clarity about whether the construct of the assessment is changed by the digital-based format. Additionally, it is important to clarify how the content is to be assessed differently using digital techniques. Although, the new platform may not substantially alter the construct being assessed, the design implications of the digital-based formats should be elaborated so that the revised framework is consistent with this new delivery system.

The Assessment Framework Development Policy does not address delivery systems or related procedural details, rather these details are addressed in procedural requirements included in framework procurements. (NAGB, 2018a, p. 19) One of the rationales for seeking framework updates going forward includes incorporating new items that will more fully capitalize on current advances in digital-based assessment. The ADC and Governing Board staff need to determine if the policy should contain guidance specifying the extent to which frameworks should include content addressing platform-specific elements. (NAGB, 2018b)

VII. Reflections and Recommendations

Reflections on Framework Development Changes

Over time, the procedures for implementing frameworks have evolved in several important ways. Beginning with the frameworks developed since the early 2000s, the frameworks and process reports have demonstrated the broad representation in this work, have included more thorough documentation of the activities conducted, and have validated the increased public comment. While the authorizing legislation and the Governing Board Framework Development Policy are important, their influence on the frameworks has not really changed. In my opinion, the law and the policy have not been the primary drivers of these changes. The greatest influencer in these changes has been the increased utilization of test information for accountability decisions and the increased expectations for test publishers, including NAEP, because of this increased use.

Broad Representation. The framework committees have always included representation of subject-area experts (academicians and curriculum specialists), educators (teachers, local and state administrators), policy makers, parents, and the general public. Additionally, they were diverse in terms of gender, ethnicity/race, region, and representation of public-private school students, high-poverty students, and low-performing school students. When the participation

of all students in NAEP and accommodations were added to the assessment, persons who specialize in assessing students with disabilities and English learners also were included. Documentation of participants in framework committees and in the public comment opportunities provides evidence of this broad representation.

More Thorough Documentation. The framework documents produced today provide much more detail than the first framework documents, especially in terms of item examples and information about achievement levels. An example is found in the 1996 and the 2019 Mathematics Assessment Frameworks for NAEP. The 1996 Mathematics Framework includes three example items, one for each type of item to be included in the assessment: multiple-choice, open-ended, and extended open-ended. In contrast, the 2019 Mathematics Framework includes 14 unique items, five to describe the types of items included in the assessment (multiple-choice, short constructed-response, and extended constructed-response), and nine to provide examples of pure mathematics items (four items), calculator involved items (three items), and items using manipulatives (two items). In addition, the 2019 Mathematics Framework included a separate discussion of accessibility to item content for students with disabilities and English learners, after the examples of items. More detailed information about item design and accommodations is found in the *Assessment and Item Specifications for the NAEP Mathematics Assessment*. (NAGB, 1992; NAGB, 2006a; NAGB, 2007)

Another example of more thorough documentation in framework documents is the description about NAEP achievement levels. The 1996 framework describes the achievement levels in a single paragraph.

> The new NAEP Mathematics Framework was considered in light of the three NAEP achievement levels basic, proficient, and advanced. These levels are intended to provide descriptions of what students should know and be able to do in mathematics. Established for the 1992 mathematics scale through a broadly inclusive process and adopted by the Governing Board, the three levels per grade are a major means of reporting NAEP data. The new mathematics assessment was constructed with these levels in mind to ensure congruence between the levels and the test content. (NAGB, 1992, p. 3)

However, the 2019 Mathematics Framework, provides much more information, including achievement level descriptions. An introduction to achievement levels and the policy definitions are provided in the overview section (page 2) and an entire appendix is devoted to the achievement level descriptions (pages 71-76). Descriptions are provided for each grade level and for each of the three levels (basic, proficient, and advanced) within each grade level. Also provided are the scale score points associated with each achievement level. A great deal of detail is provided in these descriptions; in fact, the grade twelve descriptions require three pages. (NAGB, 2006a)

Greater Visibility and Debate. The advent of reporting scores on NAEP which were associated with individual locales has been a huge driver for the visibility of and debate about what is

assessed. When the Governing Board was authorized in legislation, preparations had been made to provide an opportunity for states to participate voluntarily in NAEP and receive scores for their own state. One of the major concerns about the Trial State NAEP project was the content, or framework, for the assessment. In fact, a mathematics content committee was formed and they developed an objectives-based approach similar to what states would have used. Although NAEP had always been developed under the scrutiny of subject matter experts, this became the most visible and extensive review process for the assessment content up to that time.

The greatest visibility and debate about NAEP came as a result of the No Child Left Behind Act (NCLB) in 2001. Some states had been participating in NAEP voluntarily for several years, however NCLB required all states to participate. Further the NCLB requirements revealed that NAEP would be used to evaluate the progress being reported by states on their own state tests and based on their own proficiency definitions. The publication of state-by-state NAEP results, especially in terms of the percent proficient, became controversial and the topic of much debate. In 2003, NCES began comparing each state's standard for proficient performance in reading and mathematics at grades 4 and 8 by placing the state standards onto a common scale of the National Assessment of Educational Progress (NAEP). The periodic report, *Mapping State Proficiency Standards* onto the NAEP scales also created much discussion and debate in the educational assessment community. (NCES, 2009; Ho and Haertel, 2007a; Ho and Haertel, 2007b)

There were claims that the NAEP content was different from state content and that the levels of proficiency for NAEP were higher than typical grade level expectations for students. There was partial truth in these claims, but the claims did not acknowledge the intentional design differences between NAEP and state assessments, including the intended meaning of the achievement levels, especially proficient. From the beginning NAEP frameworks had avoided matching its framework to a single set of content objectives and had strived to be broadly representative of the content domain. The NAEP frameworks were never intended to be a curriculum framework, like the standards states use, and never claimed to be. In addition, in setting the NAEP achievement levels, the Governing Board did not want them necessarily to reflect only the current level of student achievement. The desire was to define the content students should know across a range of achievement. Therefore, educators were asked to identify content expectations for basic, proficient, and advanced levels of achievement. The debates about the use of the word “proficient” and the alignment of it with state definitions of proficiency, and the alignment of NAEP frameworks with state standards will continue as long as comparisons of results are made across different locales, different assessments, and using different performance level definitions.

Another concern about the content defined in the NAEP assessment frameworks was how to consider the impact of the Common Core State Standards and their subsequent adoption/implementation in numerous states. The National Governors Association supported this initiative and the U.S. Education Department provided grants (via several consortia projects) to support states in revising their standards and assessments to align with the

"common core." During this period, there also were calls for the NAEP frameworks to be aligned with the common core and alignment studies were conducted by groups external to the Governing Board. (Daro, et.al., 2015) Recently, comprehensive reviews of state standards were conducted by the Governing Board for mathematics and science. (AIR, 2018a, 2018b, 2018c. 2018d; HumRRO 2021) Prior to wide-spread adoption of the "common core," there was much less convergence across state standards and expectations for students. This variability had historically impacted the feasibility and understandability of studies about the relationship of NAEP to state standards.

External Input/Public Comment. Input into the first NAEP content frameworks was obtained both from the committee members who recommended the content to the Board and from individuals and national organizations external to this work. Staff solicited comments on frameworks as well as posted notices of the Board's intended actions in the Federal Register, a legal requirement still in effect. Today, proactive outreach activities for the purpose of obtaining feedback on the draft frameworks are required in the procurements issued by the Governing Board (NAGB, 2018a, p. 18). Contractors conduct these activities and document them in process reports prepared for the Governing Board. (WestEd, 2006, 2010, 2021)

The 2018 Framework Development Policy recognizes that external input is important. In fact, the policy calls for the identification of substantive issues at the beginning of the process to review the framework so these can be addressed during the project to develop or update the framework. "… the ADC shall solicit input from experts to determine if changes are warranted, making clear the potential risk of changing frameworks to trends and assessment of educational progress." (NAGB, 2018b, p. 6) Additionally, framework development project staff conduct extensive external reviews of the draft framework before a final draft is presented to the Board for adoption.

The excerpts below from the most recent process report for the NAEP Mathematics Framework illustrate the extensiveness of the outreach efforts conducted before the Board is presented a final draft for adoption. (WestEd, 2021, pp. E-3-4)

> "Outreach to organizations and individuals … was conducted with assistance from a number of collaborating organizations including the Council of Chief State School Officers (CCSSO), Conference Board for the Mathematical Sciences (CBMS) and its member organizations, National Council of Teachers of Mathematics (NCTM), TODOS: Mathematics for ALL (TODOS), Benjamin Banneker Association, National Council of Supervisors of Mathematics (NCSM), Association of Mathematics Teacher Educators (AMTE), Mathematical Association of America (MAA), and Mathematical Sciences Research Institute (MSRI).
>
> "Organizations (e.g., NCTM, AMTE, TODOS, MAA) disseminated information about the project website (naepframeworkupdate.org) and through flyers, email newsletters, social media, website announcement, hosted webinars, and

podcasts. In conjunction with partnership organizations, WestEd facilitated six live webinars, five in-person presentations, and one podcast recording.

"Across in-person and live venues, more than 1,000 people participated in outreach activities from the target stakeholder groups: Teachers, Curriculum Specialists, Content Experts, Assessment Specialists, State Administrators, Local School Administrators, Instructional Leaders, Policymakers, Business Representatives, Parents, Students, Users of Assessment Data, Researchers and Technical Experts, and other interested Members of the Public.

"Across digital communications, … email and social media dissemination of information reached more than 25,000 people across the target audiences … ."

Important Policy Updates

When the *Framework Development Policy* was revised in 2018, adding a process for updating frameworks was conceptually important. Time will tell if it is of any practical significance. The Governing Board is such a deliberative body, it is not assumed that the time for completing an update will be substantially shorter than for creating a new framework. Additionally, it is unknown how receptive the users of NAEP will be to "minor" revisions to the framework. Of course, this is both a perception and a communication challenge, and only the communication concern can be addressed by Board actions.

Removing procedures from policy is a good practice, because policy documents should provide guidance about processes and describe desirable outcomes (e.g., a valid and reliable assessment). Changes in methodology and processes should be informed as much as possible by current research and accepted best practice. If these were to become embedded in a policy, frequent revisions might be necessary and become very burdensome. A policy should focus on the big picture. The 2018 changes to the policy successfully addressed this concern.

The updates to the *Framework Development Policy* made in 2018 included: incorporating the Development Panel as part of the Visioning Panel, specifying the expected size of the panels, and utilizing technical experts in a different manner. Each of these changes are important and should facilitate the process of framework development going forward. Incorporating the Development Panel into the Visioning Panel will facilitate the ongoing work of the panelists who will be revising the framework itself. Since these panelists will have heard and participated in the discussion of issues and rationales, they should be well prepared to implement the vision for the new framework. Limiting the size of the panels will facilitate the communication of panel members with one another and be more conducive to the consensus building process. Finally, having the technical advisors available or participating in the Visioning Panel and Development Panel meetings will expedite the resolution of any technical concerns. All of these changes seem fitting and logical.

The revised 2018 *Framework Development Policy* has carefully addressed the use of classroom teaching expertise in the work of revising/updating NAEP frameworks. Almost everyone agrees that the involvement of classroom teachers is critical. That said, doing the work of revising a framework is time-consuming. Although framework projects include funds for substitute teachers' pay, it is likely that few active teachers or their administrators will be open to extended out-of-classroom time (approximately 15 days for a recent framework development process). The revised policy has addressed this tension by placing the importance on having classroom teaching experience on the Visioning Panel which requires less out-of-classroom time than the Development Panel. All members of both panels must be well qualified by content expertise and familiarity with the knowledge, skills, and abilities in the respective subject. Classroom teaching experience ensures that familiarity with the assessed grade levels will be included.

Recommendations

After reviewing mountains of minutes and many reading and mathematics framework iterations, as well as some historical documentation and reports, there are a few changes which seem worth considering.

Digital-based Assessments. Some questions in this area come to mind. Do the frameworks and specifications adopted by the Board adequately address both paper-based and digital-based assessments, especially in regard to the sample items included? Is an assessment in the digital space something about which the Governing Board needs a separate policy? A staff and committee discussion of these topics would be worthwhile.

Item Review Feedback. The Governing Board and NCES staff should discuss and develop a feedback loop process utilizing the item review standing committees. In particular, this feedback loop should focus on identifying elements in the framework that could be revised because the assessment of them lacks fidelity to the desired outcome as intended in the framework.

Continued Discussion Needed. Although the construct of "dynamic frameworks" is alluring, it has not been defined operationally in a sufficient enough manner to evaluate its practicality for the Governing Board. At this point, a recommendation for future consideration is all that can be offered. Further study and implementation details are definitely necessary to make such a proposal viable. Perhaps the standing committee feedback loop is a first step for identifying small changes that are needed in a framework to clarify how the content will be assessed.

Suggestions

The following list of suggestions are related to Framework publications. They are not presented in any order of importance and are offered for consideration of the Board and staff.

- The professional assessment standards cited in the Framework Development Policy also should be cited in framework documents because readers of these should not be left to wonder if they were utilized and implemented where applicable.
- The framework documents typically include a section of major changes. It would be helpful if these were expanded to include the rationale for the changes that were made.
- While it is important to issue framework documents corresponding to each administration of NAEP, more clarity is needed about when the Board actually adopted the framework represented in the publication. Having this embedded in the report is fine, but not sufficient for easy historical clarity. The title of the document should be augmented to contain the adoption date.
- Given the 2018 Framework Development Policy about updating frameworks, the framework document should clarify if the framework represents a major revision that may impact trend or if only minor updates were made, i.e., to incorporate digital-based items. While this is may be an empirical issue, the framework document should indicate whether special analyses will be conducted to make this determination.
- The framework documents need to include a little more about the "big picture" process followed in producing the framework, including references and links to expert testimony and public hearings which led to adoption by the Governing Board. This need not detract from the presentation of the content, but could be included as an appendix along with the names of panel members.

References

The references reviewed for this report are organized into five categories. The general category, appearing first, includes all resources that did not fall under the other titles. The other categories are: Legislation, Assessment Frameworks and Reports, National Assessment Governing Board Policies, and Governing Board Meeting Materials, Minutes and Transcripts.

General

Alexander, L., and James, H.T., (1987). *The Nation's Report Card: Improving the Assessment of Student Achievement*. (Report of the Study Group with a review of the report by a Committee of the National Academy of Education.) Washington, D.C.: National Academy of Education.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing.* Washington, D.C.

Bourque, M. L. (2004). Chapter 8: A History of the National Assessment Governing Board. In Jones, L. V., and Olkin, I. (Eds.). (2004). *The nation's report card: evolution and perspectives*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Buckendahl, C.W., Davis, S.I., Plake, B.S., Hambleton, R.K. Zenisky, A.L., and Wells, C.S. (2009). *Evaluation of the National Assessment of Education Progress: Study Reports*. Washington, DC. Department of Education.

Daro, P., Hughes, G. B., & Stancavage, F. (2015). Study of the Alignment of the 2015 NAEP Mathematics Items at Grades 4 and 8 to the Common Core State Standards (CCSS) for Mathematics. Retrieved from https://www.air.org/project/naep-validity-studies-nvs-panel

Haertel, E., Beauregard, R., Confrey, J., Gomez, L., Gong, B., Ho, A., Horwitz, P., Junker, B., Pea, R., Rothman, R., Shepard, L. (2012). *NAEP: Looking Ahead, Leading Assessment into the Future*. A panel report on ideas and recommendations for the future of NAEP. Retrieved from https://nces.ed.gov/nationsreportcard/about/future_of_naep.asp.

Ho, A. D., and Haertel, E. H., (2007a). (Over)-Interpreting mappings of state performance standards onto the NAEP scale. Paper commissioned by the Council of Chief State School Officers. Retrieved from: https://scholar.harvard.edu/andrewho/publications/over-interpreting-mappings-state-performance-standards-naep-scale

Ho, A. D., and Haertel, E. H. (2007b). Apples to apples? The underlying assumptions of state–NAEP comparisons. Paper commissioned by the Council of Chief State School Officers. Retrieved from https://scholar.harvard.edu/andrewho/files/ho_haertel_apples_to_apples.pdf

Joint Committee on Testing Practices. (2004). *Code of Fair Testing Practices in Education*. Washington, D.C.

Jones, L. V., and Olkin, I. (Eds.). (2004). *The nation's report card: evolution and perspectives*. Bloomington, IN: Phi Delta Kappa Educational Foundation in cooperation with the American Educational Research Association.

Lehmann, I. J. (2004). Chapter 1: The Genesis of NAEP. In Jones, L. V., and Olkin, I. (Eds.). (2004). *The nation's report card: evolution and perspectives*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Mullins, I.V.S. (2017) *Content of the Long-Term Trend Assessments Compared to Main NAEP*. A reaction paper prepared for the National Assessment Governing Board. Retrieved from https://www.nagb.gov/news-and-events/news-releases/2017/2017-long-term-trend-symposium.html.

National Academies of Sciences, Engineering, and Medicine. (2017). *Evaluation of the Achievement Levels for Mathematics and Reading on the National Assessment of Educational Progress*. Washington, DC: The National Academies Press. doi: 10.17226/23409.

National Center for Education Statistics-NCES. (2012). *Statistical Standards*. Retrieved from https://nces.ed.gov/statprog/2012/.

National Center for Education Statistics-NCES. (2020a). *Beginning of NAEP.* Retrieved from https://nces.ed.gov/nationsreportcard/about/newnaephistory.aspx#beginning.

National Center for Education Statistics-NCES. (2020b). *NAEP Item Development Process.* Retrieved from https://nces.ed.gov/nationsreportcard/contracts/item_dev.asp.

National Center for Education Statistics-NCES, Institute of Education Sciences. (2009). *Mapping State Proficiency Standards*. http://nces.ed.gov/nationsreportcard/studies/statemapping/.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: NCTM.

National Council of Teachers of Mathematics. (1991). *Professional Standards for Teaching Mathematics*, Reston, VA: NCTM.

National Research Council. (1999). *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress*. Committee on the Evaluation of National and State Assessments of Educational Progress. Pellegrino, J.W., Jones, L.R., Mitchell, K.J. (Eds.). Board of Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academy Press.

Selden, R. (2004). Chapter 7: Making NAEP State-by-State. In Jones, L. V., and Olkin, I. (Eds.). (2004). *The nation's report card: evolution and perspectives*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

U.S. General Accounting Office (1976). *The National Assessment of Educational Progress: Its Results Need to be Made More Useful*. HRD–76–113; Washington, DC: U.S. Government Printing Office.

Vinovskis, M.A. (1998). *Overseeing the nation's report card: the creation and evolution of the National Assessment Governing Board (NAGB)*. Paper prepared for the National Assessment Governing Board. Washington, DC. http://www.nagb.org/publications/95222.pdf.

Wirtz, W., and Lapointe, A. (1982). *Measuring the Quality of Education: A Report on Assessing Educational Progress*. Washington, D.C. (Willard Wirtz and Archie E. Lapointe)

Legislation

Public Law 100-297. (1988). *Hawkins-Stafford Elementary and Secondary School Improvement Amendments of 1988*, § 100-297 U.S.C., Title III, Part C-National Assessment of Educational Progress Improvement Act § 3403 (pp. 215-220). Retrieved from https://www.govinfo.gov/content/pkg/STATUTE-102/pdf/STATUTE-102-Pg130.pdf.

Public Law 103-382. (1994) *Improving America's Schools Act of 1994*, § 103-382 U.S.C. § 412 (pp. 522-525). Retrieved from https://www.congress.gov/public-laws/103rd-congress.

Public Law 107-110. (1994). *No Child Left Behind Act of 2002,* § 107-110 U.S.C. § 412 (pp. 1904-1907). Retrieved from https://www.govinfo.gov/app/details/PLAW-107publ110.

Public Law 107-279. (2002). *Education Sciences Reform Act of 2002*, § 107–297 U.S.C. § 302 (pp. 44-45). Retrieved from https://www.govinfo.gov/link/plaw/107/public/279.

Assessment Frameworks and Reports

American Institutes for Research-AIR (2018a). Review of State Curricular Standards in Mathematics: Executive Summary. Unpublished manuscript. Retrieved from the National Assessment Governing Board.

American Institutes for Research-AIR (2018b). Review of State Curricular Standards in Mathematics: Methodological Narrative. Unpublished manuscript. Retrieved from the National Assessment Governing Board.

American Institutes for Research-AIR (2018c). Review of State Curricular Standards in Mathematics: Study Report. Unpublished manuscript. Retrieved from the National Assessment Governing Board.

American Institutes for Research-AIR (2018d). Review of State Curricular Standards in Mathematics: Supplemental Exhibits. Unpublished manuscript. Retrieved from the National Assessment Governing Board.

Achieve, Inc. (2005). *Recommendations to the National Assessment Governing Board on Aligning 12th Grade NAEP with College and Workplace Expectations: Reading*. Retrieved from the National Assessment Governing Board.

Achieve, Inc. (2006). *Recommendations to the National Assessment Governing Board on Aligning 12th Grade NAEP with College, Workplace and Military Expectations: Mathematics*. Retrieved from the National Assessment Governing Board.

Educational Testing Service. (2010a). *Final Report on the Study to Draft Achievement-Level Descriptions for Reporting Results of the 2009 National Assessment of Educational Progress in Reading for Grades 4, 8 and 12*. Prepared under contract to and in conjunction with the National Assessment Governing Board. Retrieved from https://www.nagb.gov/focus-areas/reports.html?q0=ets.

Educational Testing Service. (2010b). *Final Report on the Study to Draft Achievement-Level Descriptions for Reporting Results of the 2009 National Assessment of Educational Progress in Mathematics for Grade*

12. Prepared under contract to and in conjunction with the National Assessment Governing Board. Retrieved from https://www.nagb.gov/focus-areas/reports.html?q0=ets.

Human Resources Research Organization-HumRRO (2021). Comparative Analysis of the NAEP Science Framework and State Science Standards: Technical Report. Unpublished manuscript. Retrieved from the National Assessment Governing Board.

Jago, Carol (2009). *A History of NAEP Assessment Frameworks.,* A paper commissioned for the 20th Anniversary of the National Assessment Governing Board 1988-2008. Retrieved from https://www.nagb.gov/news-and-events/calendar/special-events/20th-anniversary.html.

National Assessment Governing Board-NAGB. (1990). *Reading Framework for the 1992 National Assessment of Educational Progress*. Retrieved from the National Assessment Governing Board. (Note: this framework was also applicable to the assessments conducted in 1994, 1998, 2000, 2002 and 2003, although separate editions were published for each assessment year.)

National Assessment Governing Board-NAGB. (1992). *Mathematics Framework for the 1996 National Assessment of Educational Progress*. Retrieved from the Education Resources Information Clearinghouse – Document ED 404 184. (Note: This framework was also applicable to the assessments conducted in 2000 for all grades and in 2003 for grades 4 and 8, although separate editions were published for each assessment year.)

National Assessment Governing Board-NAGB. (2001a). *Mathematics Framework for the 2005 National Assessment of Educational Progress*. Retrieved from the National Assessment Governing Board. (Note: This framework was also applicable to the assessments conducted in 2007, although separate editions were published for each assessment year.)

National Assessment Governing Board-NAGB. (2004a). *Reading Framework for the 2005 National Assessment of Educational Progress*. Retrieved from the National Assessment Governing Board. (Note: This framework was also applicable to the assessment conducted in 2007, although a separate edition was published for that year.)

National Assessment Governing Board-NAGB. (2006a). *Mathematics Framework for the 2009 National Assessment of Educational Progress*. Retrieved from the National Assessment Governing Board. (Note: This framework was also applicable to the Grade 12 assessments conducted 2013 and 2016, and to the Grades 4 and 8 assessments conducted in 2011, 2013, 2015, 2017, 2019, although separate editions were published for each assessment year.)

National Assessment Governing Board-NAGB. (2006b). *Reading Framework for the 2009 National Assessment of Educational Progress*. Retrieved from https://www.nagb.gov/naep-frameworks/reading.html. (Note: This framework was also applicable to the assessments conducted in 2011, 2013, 2015, 2017 and 2019, although separate editions were published for each assessment year.)

National Assessment Governing Board-NAGB. (2007). *Assessment and Item Specifications for the NAEP 2009 Mathematics Assessment*. Retrieved from https://www.nagb.gov/naep-frameworks/mathematics.html.

National Assessment Governing Board-NAGB. (2018a). Attachment A: Statement of Work. *Update of NAEP Frameworks for Mathematics, Reading and Other Subjects (RFP# 91995918R0002)*. Retrieved from https://www.nagb.gov/governing-board/contract-opportunities.html.

National Assessment Governing Board-NAGB. (2020a). *Mathematics Framework for the 2025 National Assessment of Educational Progress*. Unpublished Manuscript. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB (2020b). *Strategic Vision 2025*. Retrieved from https://www.nagb.gov/governing-board/strategic-vision.html.

National Assessment Governing Board-NAGB. (2021). *Draft: Process Report – 2025 NAEP Mathematics Framework Update*. Unpublished Manuscript. Retrieved from the National Assessment Governing Board.

National Assessment of Educational Progress-NAEP (1988). *Mathematics Objectives: 1990 Assessment*. Princeton, N.J.: Educational Testing Service. Retrieved from the National Assessment Governing Board. (Note: this framework

National Assessment of Educational Progress-NAEP (1989). *Reading Objectives: 1990 Assessment*. Princeton, N.J.: Educational Testing Service. Retrieved from the National Assessment Governing Board.

WestEd (2006). *Final Report: NAEP Science Assessment 2009*. Retrieved from the National Assessment Governing Board.

WestEd (2010). *Final Report: Technology and Engineering Literacy Assessment for the 2014 National Assessment of Educational Progress.* Retrieved from the National Assessment Governing Board.

WestEd (2021). *Process Report: 2025 NAEP Mathematics Framework Update*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board Policies

National Assessment Governing Board-NAGB. (1996). *Redesigning the National Assessment of Educational Progress*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2001b). *NAEP and the No Child Left Behind Act*. Letter to Secretary of Education from the Board Chair, on behalf of the National Assessment Governing Board. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002a). *Framework Development – Policy Statement*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2002b). *Item Development and Review – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002c). *Long-term Trend – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002d). *Plan for Study of NAEP Sampling*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002e). *Policies and Procedures for Complaints Related to the National Assessment of Educational Progress*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002f). *Prohibition on Using NAEP to Influence State and Local Standards, Tests, and Curricula – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002g). *Public Access to Test Questions, Item Release, and Confidentiality of Data for NAEP – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002h). *Resolution on Participation of the Commonwealth of Puerto Rico in NAEP – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002i). *Review of the National Assessment of Educational Progress – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2003a). *Informing Parents About Participation in NAEP – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2004b). Resolution on the NAEP 2009 Reading Framework. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep/resolution-09.html.

National Assessment Governing Board-NAGB. (2010a, 2014). *NAEP Testing and Reporting on Students with Disabilities and English Language Learners – Policy Statement.* Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2010b). By-Laws of the National Assessment Governing Board. Retrieved from https://www.nagb.gov/governing-board/overview.html.

National Assessment Governing Board-NAGB. (2015). R*esolution on Maintaining NAEP Trends with the Transition to Digital-Based Assessments (DBA)*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2018b). *Framework Development – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2018c). Developing Student Achievement Levels for the National Assessment of Educational Progress – Policy Statement. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

Governing Board Meeting Materials, Minutes and Transcripts

National Assessment Governing Board-NAGB. (2002j). *Materials for the May 2002 Board Meeting.* Retrieved from the NAGB archive upon request.

National Assessment Governing Board-NAGB. (2002k) *Meeting Transcript from the May 17-18, 2002 Board Meeting Open Session*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2002l). *Materials for the August 2002 Board Meeting.* Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2002m). *Meeting Minutes from the August 2-3, 2002, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2002n). *Meeting transcript from the August 2-3, 2002, Board Meeting Open Session*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2002o). *Meeting Minutes from the November 15-16, 2002, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2003b). *Materials for the March 2003 Board Meeting: NAEP 2007 Reading Framework Project Update - Reading Issues Papers Synthesis*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2003c). *Meeting Minutes from the March 7-8, 2003, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2003d). *Meeting Minutes from the May 16-17, 2003, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2003e). *Meeting Minutes from the August 1-2, 2003, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2003f). *Meeting Minutes from the November 14-15, 2003, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2004c). *Meeting Minutes from the March 5-6, 2004, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2004d). Meeting transcript from the April 19, 2004, Public Forum on the Draft 2009 Reading Framework. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2004e). *Meeting transcript from the August 6, 2004, Board Meeting Open Session*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2004f). *Meeting Minutes from the November 19-20, 2004, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2005a). *Meeting Minutes from the March 4-5, 2005, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2005b). *Meeting Minutes from the May 20-21, 2005, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2005c). *Meeting Minutes from the August 5-6, 2005, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2005d). *Meeting Minutes from the November 18-19, 2005, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2006c). *Meeting Minutes from the March 3-4, 2006, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2006d). *Meeting Minutes from the May 19-20, 2006, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2006e). *Meeting Minutes from the August 4-5, 2006, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2016). Meeting minutes from the November 2, 2016 Board Meeting. Retrieved from https://www.nagb.gov/governing-board/quarterly-board-meetings/archive.html.

National Assessment Governing Board-NAGB. (2017). Meeting minutes from the March 2, 2017 Board Meeting. Retrieved from https://www.nagb.gov/governing-board/quarterly-board-meetings/archive.html.

National Assessment Governing Board-NAGB. (2018c). *Materials for the March 2018 Board Meeting: 14-framework-development-policy-revision March 2018*. Retrieved from https://www.nagb.gov/governing-board/quarterly-board-meetings/2018/2018-march.html.

Appendix A
Historical Context[7] for Framework Development of the National Assessment of Educational Progress

Dates	Historical Activities	Assessment Development
1960-70's ECS era	• The 1960s were a formative time for the development of NAEP. (NCES website: https://nces.ed.gov/nationsreportcard/about/newnaephistory.aspx#beginning) • 1964-68 – The Education Commission of the States (ECS), managed and conducted the first national assessments. They established an Exploratory Committee for the Assessment Progress in Education (ECAPE) and established a National Assessment Planning Project. • 1969 – First national assessment data collection, now known as the National Assessment of Educational Progress (NAEP), was the 1969 trial assessment of the citizenship, science, and writing performance of 17-year-old in-school students in the spring of that year. In the fall, 9- and 13-year-old students as well as out-of-school 17-year-olds were assessed. • The frameworks for the early NAEP utilized a content-by-process matrix to develop items for the assessment, most of which were released with the reporting.	The assessment was based on a content-by-process matrix set of objectives developed by representatives for the Education Commission of the States (ECS).
1976-1988 Early national assessment and NAEP era[8]	• The Comptroller General (GAO) Report, Make NAEP More Useful, was released in 1976. • The original national assessment legislation in 1978 brought changes to the oversight and organization of the assessment (now NAEP) and specified an Assessment Policy Committee of 17 members (the precursor to the National Assessment Governing Board). • A major study critical of NAEP (Wirtz & Lapointe, 1982) said NAEP was underdeveloped and underutilized, and of apparently negligible influence. • In 1983, a non-profit organization (Educational Testing Service, ETS) was selected as the NAEP Contractor and a redesigned assessment (more sophisticated sampling, scaling & analyses) was developed. • The 1986 reauthorization of the Elementary and Secondary Education Act (ESEA) included provisions for voluntary state assessments and referred to the national assessment as the National Assessment of Educational Progress, the name that continues today. It also, continued the requirement for an Assessment Policy Committee of 19 members, adding two additional members representing elementary and secondary school principals.	Because of the desire by some state members of ECS, two policy pushes changed NAEP. (1) Voluntary participation and reporting on states (2) A move to an objectives-based approach instead of the content-by-process matrix approach previously used for the assessments.

[7] A thorough examination of the establishment and early years of the National Assessment Governing Board can be found in the report, *Overseeing the Nation's Report Card: The Creation and Evolution of The National Assessment Governing Board (NAGB)*. Vinovskis, M.A. (1998). http://www.nagb.org/publications/95222.pdf.

[8] A thorough examination of the evolution of the National Assessment of Educational Progress is found in the book, The Nation's Report Card: Evolution and Perspectives (Jones & Olkin, 2004).

Appendix A Historical Context[7] for Framework Development of the National Assessment of Educational Progress		
Dates	**Historical Activities**	**Assessment Development**
	• In 1986, the Trial State Assessments were begun in cooperation with the ECS and the Southern Region Education Board (SREB). The planning for this effort was advised by a mathematics content committee which wanted to develop an objectives-based approach that could lead instruction instead of the content-by-process matrix approach previously used for the assessments.	
1988 – Present NAEP-NAGB era	• The 1988 reauthorization of ESEA & NAEP (Hawkins-Stafford Act, 1988) included provisions the establishment of a separate policy board of 24 members, the National Assessment Governing Board. The Governing Board was to be of similar composition to the Assessment Advisory Committee (specifying the additional inclusion of two curriculum specialists, a non-public educator, two governors, and an ex officio member). It also included a requirement to set feasible achievement goals – achievement levels, as they have come to be called. • The 1994 reauthorization of ESEA, Improving America's Schools Act, updated the membership of the Board to 26 by adding one more test and measurement expert and delineating the general public representatives as including two parent representatives (one additional). • The 2001 reauthorization of ESEA required state participation in NAEP Reading and Mathematics if the state received Title I funds, and called for biennial testing of Reading and Mathematics, as well as the school accountability provision known as adequate yearly progress. The content and all aspects of NAEP were now being scrutinized much more strenuously. • A 2003 authorization of the NAEP legislation provided for the voluntary inclusion of urban district level reports, included additional funding for their participation which increased from six in 2003 to 27 presently. • The 2015 reauthorization of ESEA, the Every Student Succeeds Act (ESSA), maintained the requirements for NAEP, including required state participation and biennial testing in Reading and Mathematics.	The National Assessment Governing Board was established. The 1988 legislation included provisions for trial assessments in mathematics at 8th grade (1990) and 4th and 8th grade (1992) and in reading at 4th grade (1992). The first assessment frameworks were developed for these grades/subject areas. The policy and practices for developing the NAEP Assessment Frameworks was now the responsibility of the Governing Board.

Appendix B
Governing Board Duties in Legislation Over Time

The National Assessment Governing Board was authorized by Federal legislation in 1988 and has been reauthorized twice. The duties of the National Assessment Governing Board were initially authorized in the legislation establishing the Board in 1988 and have remained quite stable throughout the periodic reauthorizations, the latest of which is P.L.107-279 (2002).[9] This law provides authorization for both the Governing Board (Section 302) and NAEP (Section 303). Appendix B presents only the Governing Board section, but does contain references to the NAEP section.

In each iteration of the law the subsections have been rearranged slightly and language was added, deleted or clarified. The requirements, however, have remained essentially the same. Two unique elements were added in 2002. The first, 302(e)(1)(D), called for an inclusive review process for the assessment that is now addressed both by a Governing Board policy (NAGB, 2002i)[10] and by the extensive external reviews conducted before each framework is adopted. The other addition, 302(e)(1)(F), provided a linkage to the NAEP section. Appendix B presents all of the legal requirements in a side-by-side arrangement. Each requirement is presented with the legal numbering used in each reauthorization and identifies changes that occurred in each revision.

Appendix B **Governing Board Duties in Legislation Over Time** (New wording is underlined. Notes in red are not included in the legislation.)		
1988 P.L. 100-297 **Sec. 3403. (6)(A)**	**1994 P.L. 103-382** **SEC. 412 (e)(1)**	**2002 P.L. 107-279** **SEC. 302. (e)(1)[11]**
6(A) In carrying out its functions under this subsection, the Board shall be responsible for-	(1) In General. -- In carrying out its functions under this section the Board shall	(1) IN GENERAL- In carrying out its functions under this section the Assessment Board shall—
(i) selecting subject areas to be assessed (consistent with paragraph (2)(A));	(A) select subject areas to be assessed (consistent with section 411(b)(1));	(A) select the subject areas to be assessed (consistent with section 303(b));

[9] The 1988 authorization, Public Law 100-297, was part of the *Hawkins-Stafford Elementary and Secondary School Improvement Amendments of 1988.* The 1994 reauthorization, Public Law 103-382, was part of the *Improving America's Schools Act of 1994*.

[10] The Governing Board policy statement, *Review of the National Assessment of Educational Progress*, adopted August 3, 2002, included six guiding principles that describe expectations for the rigorous review of the National Assessment of Educational Progress and actions of the Governing Board.

[11] Public Law 107-279, the Education Sciences Reform Act of 2002, provided amendments to the original No Child Left Behind Act of 2002, Public Law 107-110.

Appendix B Governing Board Duties in Legislation Over Time (New wording is underlined. Notes in red are not included in the legislation.)		
1988 P.L. 100-297 Sec. 3403. (6)(A)	**1994 P.L. 103-382 SEC. 412 (e)(1)**	**2002 P.L. 107-279 SEC. 302. (e)(1)[11]**
(ii) identifying appropriate achievement goals for each age and grade in each subject area to be tested under the National Assessment;	(B) develop appropriate student performance levels as provided in section 411(e);	(B) develop appropriate student achievement levels as provided in section 303(e);
(iii) developing assessment objectives; (iv) developing test specifications;	(C) develop assessment objectives and test specifications through a national consensus approach which includes the active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the public; Note: Consensus process was incorporated here from 1998 section (E).	(C) develop assessment objectives consistent with the requirements of this section and test specifications that produce an assessment that is valid and reliable, and are based on relevant widely accepted professional standards; Note: Reference to a consensus approach was moved from the NAGB, Section 302, to the NAEP Section 303(b)(3)(B)(II) but still applies to the content of NAEP for which the Board is responsible.
		(D) develop a process for review of the assessment which includes the active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the public;
(v) designing the methodology of the assessment;	(D) design the methodology of the assessment, in consultation with appropriate technical experts, including the Advisory Council established under section 407;	(E) design the methodology of the assessment to ensure that assessment items are valid and reliable, in consultation with appropriate technical experts in measurement and assessment, content and subject matter, sampling, and other technical experts who engage in large scale surveys;

Appendix B Governing Board Duties in Legislation Over Time (New wording is underlined. Notes in red are not included in the legislation.)		
1988 P.L. 100-297 Sec. 3403. (6)(A)	**1994 P.L. 103-382 SEC. 412 (e)(1)**	**2002 P.L. 107-279 SEC. 302. (e)(1)[11]**
		(F) consistent with section 303, measure student academic achievement in grades 4, 8, and 12 in the authorized academic subjects;
(vi) developing guidelines and standards for analysis plans and for reporting and disseminating results;	(E) develop guidelines and standards for analysis plans for reporting and disseminating results;	(G) develop guidelines for reporting and disseminating results; Note: 'Standards for analysis plans" was removed from this section.
(vii) developing standards and Procedures for interstate, regional and national comparisons; and	(F) develop standards and procedures for interstate, regional, and national comparisons; and	(H) develop standards and procedures for regional and national comparisons; Note: 'interstate' was removed from this section.
(viii) taking appropriate actions needed to improve the form and use of the National Assessment.	(G) take appropriate actions needed to improve the form and use of the National Assessment.	(I) take appropriate actions needed to improve the form, content, use, and reporting of results of any assessment authorized by section 303 consistent with the provisions of this section and section 303; and
		(J) plan and execute the initial public release of National Assessment of Educational Progress reports. The National Assessment of Educational Progress data shall not be released prior to the release of the reports described in subparagraph (J).

Appendix B Governing Board Duties in Legislation Over Time (New wording is underlined. Notes in red are not included in the legislation.)		
1988 P.L. 100-297 Sec. 3403. (6)(A)	**1994 P.L. 103-382 SEC. 412 (e)(1)**	**2002 P.L. 107-279 SEC. 302. (e)(1)[11]**
(B) The Board may delegate any functions described in subparagraph (A) to its staff.	(2) Delegation. -- The Board may delegate any of the Board's procedural and administrative functions to its staff.	(2) DELEGATION- The Assessment Board may delegate any of the Assessment Board's procedural and administrative functions to its staff.
(C) The Board shall have final authority on the appropriateness of cognitive items.	(3) Cognitive Items. -- The Board shall have final authority on the appropriateness of cognitive items.	(3) ALL COGNITIVE AND NONCOGNITIVE ASSESSMENT ITEMS- The Assessment Board shall have final authority on the appropriateness of all assessment items.
(D) The Board shall take steps to ensure that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias.	(4) Prohibition Against Bias. -- The Board shall take steps to ensure that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias.	(4) PROHIBITION AGAINST BIAS- The Assessment Board shall take steps to ensure that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias and are secular, neutral, and non-ideological.
(E) Each learning area assessment shall have goal statements devised through a national consensus approach, providing for active participation of teachers, curriculum specialists, local school administrators, parents and concerned members of the general public.	(5) Technical. -- In carrying out the duties required by paragraph (1), the Board may seek technical advice, as appropriate from the Commissioner and the Advisory Council on Education Statistics and other experts. Note: the stakeholder list and consensus approach were moved to Section 412 (e)(1)(C).	(5) TECHNICAL- In carrying out the duties required by paragraph (1), the Assessment Board may seek technical advice, as appropriate, from the Commissioner for Education Statistics and other experts.

Appendix B Governing Board Duties in Legislation Over Time (New wording is underlined. Notes in red are not included in the legislation.)		
1988 P.L. 100-297 Sec. 3403. (6)(A)	**1994 P.L. 103-382 SEC. 412 (e)(1)**	**2002 P.L. 107-279 SEC. 302. (e)(1)[11]**
	(6) Report. -- Not later than 90 days after an evaluation of the student performance levels under section 411(e), the Board shall make a report to the Secretary, the Committee on Education and Labor of the House of Representatives, and the Committee on Labor and Human Resources of the Senate describing the steps the Board is taking to respond to each of the recommendations contained in such evaluations.	(6) REPORT- 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. Note: This change provides an update to the House and Senate Committee names at the time.

Appendix C
Framework Development Policy Revision 2002 to 2018

The NAGB Framework Development Policy was developed initially in 2002 and revised 16 years later in 2018. The original policy was based on the accepted best practice NAGB had been following since 1988. Although many changes occurred in assessment methodologies and education policy, the 2002 policy served the Board will, even with some redundancies and procedural details not usually found in policies. Revisions to the Framework Development Policy in 2018 addressed these issues.

In addition to some minor reorganization and rewording, primary distinctions between the 2002 and 2018 editions included four changes that are discussed in more detail within this report: (1) updating frameworks, (2) reviewing frameworks, (3) participants/stakeholders, and (4) framework panels/committees. Additionally, the current policy maintains a focus on the overarching principles to be followed, with the details and procedures moved to procedural documents and requirements for contractors.

Basically, the two versions address the same content, although they are arranged somewhat differently and with fewer procedural elements in 2018. The summary below compares the principles in each version, in a side-by-side manner, and summarizes the changes that were implemented in 2018 (shown in red). Italicized words show *2002 language that was changed* and underlining shows new wording in 2018. Of course, this summary does not capture all changes as the text under each principle also was revised in a similar manner to remove redundancy and procedures, and for more clarity and efficiency in wording. A few are noted in the table. The only substantive change is the addition of a framework update process which is not intended to be as extensive as the development of a new framework.

Policy Elements	5/18/02 Framework Development Policy	03/18/18 Framework Development Policy
Preface: Purpose	It is the policy of the National Assessment Governing Board to conduct a comprehensive, inclusive, and deliberative process to determine the content and format of all subject area assessments under the National Assessment of Educational Progress (NAEP).	No change
Preface: Desired Outcome	Objectives developed and adopted by the Governing Board as a result of this process shall be used to produce NAEP assessments that are valid and reliable, and that are based on widely accepted professional standards. *The process shall include the active participation of educators, parents, and members of*	The primary result of this process shall be an assessment framework (hereafter, "framework") with objectives to guide development of NAEP assessments for students in grades 4, 8, and 12 that are valid, reliable, and reflective of widely accepted professional standards.

Policy Elements	5/18/02 Framework Development Policy	03/18/18 Framework Development Policy
	the general public. The primary result of this process shall be an assessment framework to guide NAEP development at grades 4, 8, and 12	Rewording & reorganization of italicized details
Preface: Process	*The process shall include the active participation of educators, parents, and members of the general public.*	This process detail is contained in the introduction and in Principle 2
Preface: Board Delegation to ADC	The Governing Board, through its Assessment Development Committee, shall *carefully* monitor the framework development process to ensure *that all Governing Board policies are followed; that the process is comprehensive, inclusive, and deliberative; and* that the final Governing Board-adopted framework, specifications, and background variables documents *are congruent with the Guiding Principles, Policies, and Procedures that follow.*	The Governing Board, through its Assessment Development Committee, shall monitor the framework development and update processes to ensure that the final Governing Board-adopted framework, specifications, contextual variables documents, and their development processes comply with all principles and guidelines of the Governing Board Framework Development Policy. Rewording, reorganization of italicized details
Intro: Legal Authorization	P.L. 107-279 Section 302(e)(1) and Restatement of law requirements	No change in citation, but requirements not explicitly listed
Intro: Involvement of Stakeholders	Stakeholders were given in the restatement of the law	Expanded description of compliance with the law and identification of specific stakeholders
Intro: Professional Standards	Adherence to standards acknowledged with current publications cited.	No change except for the editions cited
The Principles	Seven (7) principles included with *policies and procedures* for implementing each. Order is shown in relation to the 2018 policy.	Six (6) principles included with guidelines for implementation. Essentially the same principles and guidelines as in 2002 (with some combining and rewording), titles were added to each principle.
	1. The Governing Board is responsible for developing an *assessment* framework for each NAEP subject area. The framework shall define the scope of the domain to be measured by delineating the knowledge and skills to be tested at each grade, the format of the NAEP assessment, and preliminary achievement level descriptions. *5. Through the framework development process, preliminary*	1. Elements of Frameworks: The Governing Board is responsible for developing a framework for each NAEP assessment. The framework shall define the scope of the domain to be measured by delineating the knowledge and skills to be tested at each grade, the format of the NAEP assessment, and the achievement levels. Define what will be tested and how, as well as

Policy Elements	5/18/02 Framework Development Policy	03/18/18 Framework Development Policy
	achievement level descriptions shall be created for each grade being tested. These preliminary descriptions shall be an important consideration in the item development process and will be used to begin the achievement level setting process.	how much students should know at each achievement level. 2002 Principle 5 incorporated with this principle
	2. The Governing Board shall develop an assessment framework through a comprehensive, inclusive, and deliberative process that involves the active participation of *teachers, curriculum specialists, local school administrators, parents, and members of the public*. (Note: This 2002 principle contained guidelines for panel members which did not explicitly require classroom experience for the subject area. *"At least 30 percent of this committee shall be composed of users and consumers in the subject area under consideration.")*	2. Development and Update Process: The Governing Board shall develop and update frameworks through a comprehensive, inclusive, and deliberative process that involves active participation of stakeholders. Addition of 'update'; redundancy in wording reduced; and move of stakeholders list to the introduction This principle more clearly identified the various panels, their purposes, shared membership expectation, classroom teaching experience (20%) in the subject area, and expected discussions about the impact on trend reporting when content changes.
	7. *NAEP assessment frameworks and test specifications generally shall remain stable for at least 10 years*.	3. Framework Review: Reviews of existing frameworks shall determine whether an update is needed to continue valid and reliable measurement of the content and cognitive processes reflected in evolving expectations of students. The addition of this principle provides an emphasis on the work of reviewing/updating frameworks and contains guidelines about reviewing/updating frameworks at least once every 10 years.
	3. The framework development process shall take into account state and local curricula and assessments, widely accepted professional standards, exemplary research, international standards and	4. Resources for the Process: Framework development and update processes shall take into account state and local curricula and assessments, widely accepted professional standards, exemplary research, international standards

Policy Elements	5/18/02 Framework Development Policy	03/18/18 Framework Development Policy
	assessments, and other pertinent factors and information.	and assessments, and other pertinent factors and information. Addition of 'update' This principle contains expanded guidance on ways to identify curricular content issues in the field.
	6. The specifications document shall be developed *during the framework process* for use by NCES *and the test development contractor* as the blueprint for constructing the NAEP assessment and items *in a given subject area.*	5. Elements of Specifications: The specifications document shall be developed for use by NCES as the blueprint for constructing the NAEP assessment and items. Reduce unnecessary words
	4. The Governing Board, through its Assessment Development Committee, shall *closely* monitor all *steps in the* framework development *process*. The result of this process shall be recommendations for Governing Board action in the form of three key documents: the assessment framework; assessment and item specifications; and *background* variables that relate to the subject being assessed.	6. Role of the Governing Board The Governing Board, through its Assessment Development Committee, shall monitor all framework development and updates. The result of this process shall be recommendations for Governing Board action in the form of three key documents: the framework; assessment and item specifications; and contextual variables that relate to the subject being assessed. Addition of 'update' & change of term from background to contextual variables. This principle contains guidelines about balancing the maintenance of trends with including new content.

Appendix D
Decision Points and Roles for Framework Development

Appendix D highlights the major questions/decisions and other subordinate ones needed for framework development, approval, and adoption by the Board. Also included are the likely roles and involvement of contractors and external reviewers, that is, stakeholders and the general public. Many smaller decisions and steps are behind these major decision points, but cannot be captured in this simplistic presentation. While the decision points are presented in an orderly manner, they may not always be implemented in the chronology implied by this list.

Appendix D Decision Points and Roles for Framework Development				
Activity	**Full Board**	**Assessment Development Committee***	**Contractor Activities**	**External Reviews**
① Should a framework revision or updating be considered?		- Identify need for review - Recommend going forward with review		
Experts make presentations to the Assessment Development Committee.		- Convene experts - Review relevant research		
Formulate a recommendation about update/replacement of framework and draft charge		- Formulate recommendation - Draft charge		
② Is a new framework or update needed?	**Review-Approve charge**			Via public information and open meetings
Conduct procurement and select contractor to manage workload		- Issue procurement - Review proposals - Initiate Contract - Monitor*	- Begin contract and implement as required	Via public postings and notices
<u>Visioning Panel Deliberations</u> (includes Development Panel members) <u>Purpose</u>: to provide the initial high-level guidance about the state of the discipline and recommendations (guidelines or goals) for developing the framework		- Review/approve panels - Provide charge & direction - Review guidelines and goals - Regularly monitors progress*	- Identify panel chair & participants - Facilitate Process - Regularly reports progress	

Appendix D Decision Points and Roles for Framework Development				
Activity	**Full Board**	**Assessment Development Committee***	**Contractor Activities**	**External Reviews**
Development Panel Deliberations (overlap with Visioning Panel) Purpose: to draft the three project documents, engage in the detailed deliberations about how issues outlined by the Visioning Panel should be reflected in the framework		- Regularly monitors progress*	- Identify panel chair & participants - Facilitate Process - - Regularly reports progress	
Technical Experts Involved Purpose: to uphold the highest technical standards and as a resource to the framework panels to respond to technical issues raised during panel deliberations.		- Participate as needed* - Regularly monitors progress	- Identify participants - Facilitate Process - Produce Reports	
③ Is the draft framework ready to be evaluated by external reviewers? Public comment will be sought from various segments of the population to reflect many different views, and targeted feedback will be solicited from those employed in the content area under consideration, especially educators and policy makers.		- Regularly monitors progress* - Recommend going forward with external review and public comment	- Provide drafts & make revisions - Produce Reports	Via public information and open meetings
Framework – Define what, how and how much of the content domain is to be included on the NAEP assessment, and desirable levels of achievement		- Monitor* - Approve	- Facilitate Process - Produce Reports	
④-What feedback should be incorporated in Framework? The Framework Development Project must consider the policy impact and provide advice about changes needed based on the feedback, weighing all of the issues.		- Recommend activities -- Participate in activities - Review feedback - Recommend next steps	- Identify participants - Facilitate Process -- Incorporate feedback - Produce Reports	Provide verbal and written comments about the framework & other issues

Appendix D Decision Points and Roles for Framework Development				
Activity	**Full Board**	**Assessment Development Committee***	**Contractor Activities**	**External Reviews**
⑤ Should the framework be adopted and implemented? After considering the revisions made to the framework, the Board formally adopts the framework and approves the next steps.	**- Review** **- Approve or modify**	- Recommend adoption - Identify next steps (item specification and contextual variables)		
5.2 (Later) Item specifications – the blueprint for constructing the NAEP assessment in sufficient detail for developing high-quality questions based on the framework	**- Review** **- Approve or modify**	- Monitor* - Approve	- Facilitate Process - Produce Reports	
5.2 (Later) Contextual variables **–** recommendations on related contextual variables to be collected from students, teachers, and school administrators	**- Review** **- Approve or modify**	- Monitor* - Approve	- Facilitate Process - Produce Reports	
Implement Assessment in collaboration with NCES.		- Monitor* - Approve items	NCES contractors	

* Although the Assessment Development Committee has the primary role for oversight of framework development/updating processes, other committees of the Board and NCES are involved as needed. Typically, the Committee on Standards, Design, and Methodology (COSDAM) is involved in technical issues (scoring, scaling, trend reporting, etc.), the Reporting and Dissemination Committee (R & D) is involved in discussions about reporting and contextual data collection, and the National Center for Education Statistics (NCES) is involved in issues related to item development, test construction, test scoring, data analysis, and reporting.

TABLE OF CONTENTS

2021 No. 022

Assessment Framework Development Processes

Final Report

Prepared for: National Assessment Governing Board
800 North Capitol Street N.W., Suite 825
Washington DC 20002

Prepared under: ED-NAG-17-C-0002

Authors: Will Lorié
Brian Gong
Center for Assessment

Date: March 31, 2021

Assessment Framework Development Processes

Table of Contents

List of Tables

Assessment Framework Development Processes

Executive Summary

By describing what is to be assessed and how to assess it, assessment frameworks play a pivotal role in testing programs. In February 2021, the National Assessment Governing Board (Governing Board), which oversees the National Assessment of Educational Progress (NAEP), invited a technical memo to discuss the processes that large-scale assessment sponsors initiate, conduct, or commission to develop, review, or update assessment frameworks. The Governing Board was particularly interested in how the framework processes of other large-scale assessment programs and framework process standards/best practices might inform the framework processes for the NAEP.

In this technical memo, we present an organizer that enumerates the elements of assessment processes. These elements and their components classify all the decisions relevant to shaping framework processes. We developed the organizer while reviewing framework process-relevant documents for NAEP and other testing programs, such as assessment frameworks themselves, technical reports, and process reports.

Although there are no recognized standards for framework processes, we also reviewed standards or other widely consulted sources that might address aspects of framework processes, such as the *Standards for educational and psychological testing* (AERA/APA/NCME, 2014). Apart from documenting what is available regarding framework process best practices, this review informed the organizer.

Our review has two significant implications for NAEP and similar large-scale testing programs. The elements of framework processes imply a set of options that will substantially shape framework processes for a program, the resulting framework, and ultimately the resulting assessment. Assessment sponsors can make choices concerning these options, delegate those choices, or a combination.

We conclude that a sound principle of best practice in this area is for test sponsors to be aware of the framework process elements/components and their associated options. Moreover, test sponsors should be deliberate in their specification of requirements. They should provide a rationale for their choices.

A second implication is that much of the quality of the framework product depends upon the *process* used to develop the framework. Because there are few established criteria to evaluate the quality of assessment frameworks, it becomes more essential that the processes be specified well and carried out well. Programs should document, evaluate, and try to improve their framework development processes.

For NAEP and all the programs reviewed, this takes on greater importance when multiple assessment frameworks are developed and there is a desire to have similar features, specificity, and/or process quality across frameworks. Consistency in product and/or process will be a matter of deliberate design and careful implementation.

We end with seven recommendations regarding further work in this area. They include investigations of:

1. The structure of domain descriptions across different assessment frameworks.
2. The different kinds of sources informing assessment frameworks.
3. The structure of assessment objectives across different assessment frameworks.
4. Different approaches to ensuring curriculum neutrality in assessment framework development.
5. The scope of the assessment design component across different assessment frameworks.
6. Best practices for implementation fidelity evaluation and documentation for group-based processes.
7. Best practices in effective committee work, especially processes for generating, discussing, and resolving issues.

Background and Approach

Assessment Frameworks

Every modern assessment program has some definition of the intended construct to be measured, including a definition of the domain. That is typically referred to as the content framework. In addition, there will be a specification of what and how to assess to produce sufficient evidence to support the intended assessment interpretations and uses. That is typically referred to as test specifications or the test blueprint. In the NAEP program, an "assessment framework" is produced that combines definition of the content and the essential assessment specifications. The assessment framework is produced under the direction of the Governing Board, typically by committees of persons with desired expertise. The assessment frameworks specify the basic architecture of the assessment to be developed.

Statement of Work

The Center and the Governing Board developed the following statement of work at the outset of the program. It is presented here without edits.

The National Assessment Governing Board (Governing Board) invited a paper to discuss how framework/standards development processes are conducted to specify the content to be covered in an assessment (hereafter, noted as "framework processes"). In consultation with HumRRO and the National Center for the Improvement of Educational Assessment (Center), Governing Board agreed that the paper should:

1. Summarize elements of framework processes for state, national, and international assessments.
2. Compare these framework processes, articulating similarities and differences.
3. List and describe common practices for developing frameworks.

4. Evaluate which practices are appropriate for NAEP's legislative mandates, e.g., curricular-neutrality, pedagogical-neutrality, etc.

5. Describe how current NAEP framework processes reflect or do not reflect these NAEP-appropriate practices.

6. Recommend possible additional work to inform Board considerations.

Approach

To accomplish the six goals of this paper as delineated in the statement of work, we began by reviewing initial documentation provided by Governing Board. Next, we read assessment frameworks and related documentation for selected assessment programs. A set of guiding questions (presented below) informed our reading.

We selected assessment programs based on their potential relevance to the NAEP context, which assesses achievement of students' domain-specific knowledge and skills across populations governed by different educational standards or curricula.

Next, we discussed dimensions that can describe different framework process choices and their interrelationships across assessment programs. Then, we created an organizer for these choices. In the process, we proposed working definitions of key terms.

We posit that assessment program sponsors should make conscious choices concerning these features. NAEP's mandates and traditions have implications for these choices, especially when compared to other programs' framework processes. Our recommendations build upon these implications.

Scope of the Review of Framework Processes

Our review of framework processes is limited to large-scale content area-based or skills-based assessments in K-12, with mandates issued by national, (U.S.) state, or international agencies. We focused on relatively recent assessment programs (or the most recent framework processes of those programs) with publicly available documentation. We shared a list of programs to review with the Governing Board early in the project through an annotated outline. Our list is presented here as originally communicated to the Governing Board:

- NAEP
- A national assessment operating in a setting where there is a national curriculum, such as the U.K.
- A national assessment operating in a multi-curricular setting like the U.S. (if there is one)
- SAT
- ACT
- An assessment for states responding to a multi-state or national-level consensus, e.g., Common Core State Standards (CCSS)-based or Next Generation Science Standards (NGSS)-based content standards for assessment

- A non-consortium state assessment example where the state developed content standards and explicitly did not substantially adopt a widely used set of content standards
- A potential state example operating under very different constraints
- Two leading international assessment programs operating under very different conceptual relationships to curriculum
  - Programme for International Student Assessment (PISA)
  - Trends in Mathematics and Science Study (TIMSS)

We subsequently identified a non-U.S.-based national program operating in a multi-curricular setting like the U.S., with the relevant documentation publicly available. This program is the Pan-Canadian Assessment Program (PCAP). Because of the similarity between the PCAP and NAEP contexts, we conducted a relatively more thorough review of PCAP and included that review as a case study in an appendix.

We did not locate a NAEP-like program in the U.K. We subsequently reconsidered the relevance of national assessment programs in countries where there is a national curriculum.[1] Our final list excluded state testing programs that develop their own standards outside the context of a consortium. In general, state testing programs do not report much about the processes they use to derive their assessment frameworks. A useful proxy may be how state curriculum or academic content standards are developed and adopted. A review of these, however, was beyond the scope of this technical memo.

Guiding Questions for Review of Framework Processes

The following questions guided our review of framework processes for NAEP and other programs.

1. What documentation is publicly available concerning framework processes for large-scale assessments, and how thoroughly does it describe those processes?
2. What are the different legislative or other mandates for framework processes, and what do these directly or indirectly imply about those processes?
3. What are the processes for selecting steering group members and authors of assessment frameworks?
4. What are the processes for securing internal agreement during authorship, and how is dissent managed?
5. What are the parameters governing review by stakeholders or other constituencies, and how are differences of opinion managed in the review process?
6. What standards or other external guidance, if any, are referenced or consulted to guide framework processes?

[1] This is why, for example, we did not investigate Australia's National Assessment Program – Literacy and Numeracy (NAPLAN). Australia has a national curriculum and so NAPLAN would not have to contend with curricular neutrality in the same way as NAEP.

7. What are common features of framework processes across all programs, and what appears to be unique to programs or programs with specific characteristics?

8. Which features of framework processes seem most appropriate to those assessment programs with a legislative mandate similar to NAEP?

9. To what extent have NAEP framework processes reflected those features?

Definitions

The language associated with framework development processes are not often very precise, therefore we articulate some working definitions below: An *assessment framework* is a document or set of documents containing (at minimum) an assessment-oriented description of the domain assessed. A domain description is ass*essment-oriented* if it can guide assessment developers to produce assessment blueprints, item and test specifications, and similar intermediate products of assessment development. An assessment framework may also contain descriptions of construct claims (such as achievement level descriptions), specific assessment design elements (such as blueprints or acceptable item formats), and process documentation (a report of how the framework was developed). Frameworks typically also include special requirements, constraints, or criteria. (See also Martineau, Dadey, & Marion, 2018, p. 4).

A framework process is a process that results in either an approved assessment framework, an update or revision to a framework, or a decision to revise, replace, or leave a framework in place. Thus, for example, a framework process might be instantiated to determine to what extent a framework is still relevant.

An *element of a framework process* is a significant dimension of a framework process. We derived a list of elements after reviewing several assessment frameworks and related documents. We identified six elements: Initiating conditions, work product, work process, owner, timeframe, and approval.

A *specification of requirements* is a document (or a part of one) that states at least one constraint or requirement of at least one element of a framework process. By contrast, elements of framework processes may be *reported* with or without reference to any requirements. A hypothetical example of a requirements specification, which might be found in a statement of work, “The framework must include four achievement levels with descriptions of what students know and can do at the upper three levels.”

Mandate is an overarching term that covers laws, memorandums of understanding, charters, and other agreements. Even though we classify mandates as “documents,” a mandate may be verbal – for example, a charge delivered by an authority to a group in person counts as a mandate. A mandate does not have to be “documented.” A hypothetical example of an undocumented mandate is a program sponsor telling a working group to prioritize content standards above studies of how content is actually taught, assuming this instruction does not make it into any document.

Methodology

Overview of Methodology

Our goal was to develop an organizer to describe framework processes. We proceeded by reviewing the initial (NAEP) documentation provided by the Governing Board. We discussed internally salient dimensions or aspects of these processes, compared to what we knew of framework processes from other assessment programs. We drew up a list of programs to review and then scanned available documentation for references to framework processes. We continued to refine our articulations of the general "elements" of framework processes, developing some definitions to guide our approach. We did an in-depth review of one additional assessment program, after which we finalized our organizer. Finally, we collated and summarized what we could find concerning professional standards for framework processes.

Initial Documentation

We received documentation relevant to NAEP framework processes at the outset of this project. These documents include the NAEP law, NAEP's framework development policy statement, select NAEP frameworks, design documents, schedules, and studies relevant to framework processes. These documents are listed in References and Appendix A and are denoted by a single asterisk.

Rationale for Selection of Assessment Programs to Review

We looked at assessments operating at national, state, and international levels. Our goal was to select assessment programs with contexts like NAEP. Specifically, we sought out achievement assessment programs where test-takers learn through different curricula and possibly under educational authorities with varying content standards.

There are two major programs with these characteristics at the international level – the Programme for International Student Assessment (PISA) and Trends in Mathematics and Science Study (TIMSS). At the national level outside of the U.S., we discovered one other national assessment program operating in contexts like NAEP. This is the Pan-Canadian Assessment Program (PCAP). At the national level within the U.S., the ACT and SAT are the prime candidates. Finally, at the state level, there are at least as many testing programs as states. We chose to focus on processes for developing consortium-based frameworks because states otherwise rely on their own academic content standards, which inform both assessment and instruction. That context differs from NAEP, which cannot make explicit connections to instruction.

Additional Documentation Reviewed

We reviewed additional documentation from other assessment programs. There are two kinds of documents: (1) documents that may *specify requirements* for elements of framework processes, *report* them, or both; and (2) documents that purport to address standards and best practices for the elements of framework processes.

The difference between *specifying requirements for* a framework process and reporting an element of a framework process is that the former states, for example, how the framework should be structured or how the product should unfold.

The difference between a document specifying requirements and a document purporting to address standards is that the first is typically written by a test sponsor and outlines what they want the product to contain and how the process should unfold. The second type of document would include principles or guidance that should apply to *every* framework process, regardless of sponsor.

Table 1. Documents Addressing Framework Processes

Documents *specifying requirements for* or *reporting elements of* framework processes	Documents addressing or potentially addressing standards or best practices
• Mandates (Laws, memorandums of understanding, charters, and other agreements – see definitions) • Statements of work • Work plans • Assessment frameworks • Reports • Communiques • Other (websites, presentations, briefs, etc.)	• Standards • Guidelines • Assessment frameworks • Reports • Communiques • Other (websites, presentations, briefs, etc.)

We present a complete list of specific documents reviewed for this technical memo in References and Appendix A. The double-asterisked references are relevant to our review of the Pan-Canadian Assessment Program (PCAP), the closest comparison to a NAEP-like program that we could find.

Organizer: Elements of Framework Processes

We developed the following organizer during our review of framework processes for NAEP and other assessment programs. We employ the highlighted terms in the manner defined in the section on working definitions. Developing, reviewing, or updating an *assessment framework* (the "work") implies the following *elements of framework processes*. A potential source of confusion is that *work process* is an element of framework processes. "Framework processes" is an over-arching term for the many aspects of developing an assessment framework.

Note that both "work product" and "work process" are considered elements of framework processes. The first addresses the critical questions about what gets included in a framework document. One way framework documents differ is how far they go in addressing test design, for example. Broadly speaking, deciding what is in the framework document and how it should be organized is a framework process. In contrast, the second element – "work process" – is about the steps to follow to produce the framework document. These two elements are independent: It is possible for test sponsors to specify requirements for components either, neither, both.

Table 2. Framework Processes Key Components and Questions Addressed by Element

Element	Key Components*	Questions addressed**
A. Initiating conditions	None	Under what conditions will this work be initiated?
B. Work product	None	What are to be the components of the final work product?
B. Work product	Domain description	What is to be the format of an assessment-oriented description of this domain?
B. Work product	Descriptions of achievement levels	What claims about student knowledge or ability are intended?
B. Work product	Assessment design	What aspects of assessment design are to be included in the work product?
B. Work product	Documentation of process	How much of the process for producing the work product is to be included in the work product itself?
B. Work product	Basis for decision to revise/retain	In the case of a review, what is to be the basis for revising or retaining an existing framework?
B. Work product	Special requirements, constraints, and criteria	What additional requirements or constraints must be reflected in the final work product?
C. Work process	None	What is the process to be followed in producing the work product?
C. Work process	Commissioning procedures	How will a contractor be selected to produce the work?
C. Work process	Selection of authors, consultants, and working groups	How will authors, consultants, etc. be selected by the contractor?
C. Work process	Timelines and milestones	What is the timeline for the work and milestones (if any milestones)?
C. Work process	Sources informing framework; their role in the work	What other sources should inform the framework, and in what way?
C. Work process	Reconciliation	What will be the process for addressing competing views on the domain or competing requirements, such as fidelity to the domain and practical assessment constraints?
C. Work process	Internal drafting and review	What will be the process for drafting the work product? Who is to be responsible? How is internal review to be managed?
C. Work process	Role of external consultants and owners in shaping the work	How will external expertise be solicited, and from whom? How will sponsors/owners provide input, if at all, prior to work product finalization? How will feedback from these parties be incorporated?
C. Work process	External review, response, and finalization	How will external (including constituency) review be conducted? How will input from the parties be responded to? What is the process for incorporating that input into the final work product?
C. Work process	Documentation requirements	What is to be documented about the work process components?

Table 3. Framework Processes Key Components and Questions Addressed by Element (Continued)

Element	Key Components*	Questions addressed**
D. Owner	None	Who is the client or sponsor of the work product?
E. Timeframe	None	What is the timeframe for producing the work product?
F. Approval	None	What is to be the process for approving the work product?
F. Approval	Approving party	Who will be approving the work product?
F. Approval	Decision process	By what process will the work product be approved (or not)?
F. Approval	Criteria for judging the work product and process	What will be the criteria for judging the quality of the work product and process?
F. Approval	Contingencies	What procedures will be followed if the work is not approved?

Note: **Please note that a component is a subdivision of an element. *The questions are written in a format anticipating *requirement specifications* for that element or component. They could also be written to anticipate *reporting* of that element or component.

Key Aspects of Framework Processes Relevant to NAEP

Several key aspects of framework processes are particularly relevant to a large-scale assessment such as NAEP.

Table 4. Key Aspects of Framework Processes Relevant to NAEP

Key aspect of framework process	Relevant framework process elements	Documents typically specifying (S) or reporting (R) this aspect
The authority or legislative mandate for developing an assessment framework	Mandates can address all framework process elements	Mandates (S)
Framework derivation*– i.e., a description of how, given authority, legislative mandate, sources, or working groups, a person or group should derive (or derived) the assessment frameworks.	C** – The process to follow/all components	Mandates (S) Statements of work (S) Frameworks (R)
Intended relationship to academic standards or curricula of the assessed population	C – The process to follow/Sources informing the framework, and their role in the work	Mandates (S) Statements of work (S) Frameworks (R)
Intended role of standards/curricula of the assessed population	C – The process to follow/Sources	Mandates (S) Statements of work (S) Frameworks (R)
Role of education research in the content area	C – The process to follow/Sources	Statements of work (S) Frameworks (R)
Role of other frameworks	C – The process to follow/Sources	Statements of work (S) Frameworks (R)
Articulating the dividing line between the aspects of test design to be covered in the framework, from those that will be in other documents, such as test or item specifications	B – Work product/Assessment design	Statements of work (S)
Sources for the assessment design	C – The process to follow/Sources	Statement of work (S) Frameworks (R)
Authorship of framework documents	*Who authors?* is addressed in C – The process to follow/Selection of authors *How?* is addressed under the same element/Reconciliation; Internal drafting and review; External review, response, and finalization	Statements of work (S) Frameworks (R)

Notes: **Derivation of a framework means developing a new framework or reviewing an existing framework and, if applicable, revising/updating that framework. *Letters refer to labels for elements in the organizer. The format in this column is “label -element / component.”

Descriptions of Assessment Programs Reviewed

The descriptions below focus on the programs' relation to the assessed population's curricula or content standards and the extent of available documentation relevant to framework processes. We describe who is involved in drafting frameworks to the extent that such information is publicly available.

National Assessments

National Assessment of Education Progress (NAEP)

Of the programs reviewed, the National Assessment of Educational Progress (NAEP) has the most extensive documentation of framework processes.

Initiating Conditions

Conditions for initiating a particular NAEP program's framework process are not specified in the National Assessment of Educational Progress Authorization Act of 2002 ("NAEP law"). Principle 3 of the NAEP Framework Development Policy Statement ("NAEP framework policy", Governing Board, 2018), however, notes that:

> "At least once every 10 years, the Governing Board, through its Assessment Development Committee (ADC), shall review the relevance of assessments and their underlying frameworks. […] Within the 10 year period for an ADC review, major changes in the states' or nation's educational system may occur that relate to one or more NAEP frameworks. In this instance, the ADC will determine whether and how changing conditions warrant an update […]" (p. 6)

As part of our review, the Governing Board responded to the question "What triggers a framework review?" with "[F]ramework reviews often occur when there are major developments in the field, developments that need to be incorporated into the assessment. Major consensus reports from groups such as the National Academies may prompt Board discussion, etc." [personal communication (email) February 16, 2021].

While this places a timeframe within which a review must occur, it underspecifies the conditions for timing such a review.

Work Product

The NAEP framework policy specifies several components of the framework process element *work product*. If framework processes are treated broadly to include the development of test specifications, then Principle 5 (Element of Specifications) specifies aspects of the "Assessment design" component of the *work product*. Principle 1 (Elements of Frameworks) explains that the frameworks should contain a description of the domain.

However, the NAEP framework policy does not specify how descriptions should be formatted or structured to fit within specific measurement paradigms – for example, it might be an implicit requirement that items must be nested within the smallest units of the framework and that tests should conform to unidimensional IRT with 3-5 major groupings of items.[2] NAEP framework

[2] This is only an example, not a recommendation from the authors.

policy Principle 5, Guideline (c), implies that the framework should have "content" and "process" dimensions.

Some components of the *work product* are further specified in NAEP framework revision statements of work, such as that attached to RFP# 91995918R0002 (Governing Board, 2018).

Work Process

As with the *work product*, the NAEP framework policy addresses several components of the framework process element *work process*. Principles 2 (Development and Update Process), 3 (Framework Review), and 4 (Resources for the Process) all address *work process* components. Two Guidelines, (b) and (d), under Principle 6 (Role of the Governing Board), also address the *work process*.

In general, the NAEP framework policy guidelines provide parameters for the components of processes but do not specify them. For example, Principle 2 highlights the need to represent a variety of viewpoints regarding the content of the assessment. However, the NAEP framework policy does not prescribe a panel-selection process to ensure this outcome. This leaves open the question of how the panel selection process should actively include those who hold minority or less popular views on the content assessed. The same applies to the framework review guidelines under Principle 3. The choice of experts from whom the Assessment Development Committee (ADC) is to solicit input can make a difference in determining whether changes are warranted, as there are often significant differences of opinion among experts. These considerations pertain to the *work process* component "Selection of authors, consultants, and working groups."

Guideline (f) of Principle 2 indicates that "protocols shall be established to support panel deliberations and to develop a unified proposal for the content and design of the assessment." (p. 6) A critical component left unaddressed at the NAEP-wide level is the process by which differences will be resolved to move forward in case consensus is not reached, called "Reconciliation" in the organizer.

A recent NAEP design document lays out a three-step approach to reconciliation, which might serve as a starting point for a cross-program reconciliation protocol:

> The first strategy will involve a process for reconciling differences in points of view relevant to the assessment framework. An overview of panel norms will be presented at the Visioning Panel meeting, with emphasis placed on building consensus. The second strategy will include a process to follow when agreement cannot be reached. For example, when the Development Panel cannot agree, it will define and document the contentious issues and differences that cannot be reconciled. If differences are technical and related to measurement, the issues will be brought to the TAC [Technical Advisory Committee]. Other issues will be sent to the project expert advisory group, who will consider the arguments and provide advice on reconciliation. If, after consulting with the TAC and/or advisory group, differences persist, the Development Panel will generate alternative options with the pros and cons articulated and priorities suggested, which can be reviewed during the public comment phase of the project. (WestEd, 2019, pp. 14-15)

(Note that reconciliation protocols should anticipate potentially unreconcilable differences of opinion at every stage where multiple individuals, including experts and the public, provide input or feedback.)

For NAEP, the *work product* includes descriptions of achievement levels (ALDs). Principle 1 of the NAEP framework policy indicates that framework development entails answering "how much" of content domain students should know and be able to do at the three NAEP levels. Still, aside from needing to be based on the Governing Board's very general policy definitions, there is little guidance on how to derive these descriptions. The Governing Board's Policy on achievement levels (Governing Board, 2018) explains that achievement levels consist of three parts: ALDs, cut scores, and exemplar items or tasks. That policy indicates early in the document that the development of ALDs "shall be completed initially through the process that develops the assessment frameworks." (p. 5). The remainder of the Policy on Achievement Levels appears to focus on standard setting, a process into which ALDs serve as *input*. The NAEP framework policy does not specify a process for developing ALDs.

The NAEP framework policy partially addresses the *work process* component "Sources informing the framework, and their role in the work" under Principle 4 (Resources for the Process). Several resources are mentioned, including:

> An initial compilation of resources" that "summarize[s] relevant research, advantages and disadvantages and latest developments, and trends in state standards and assessments in the content area. […And] curriculum guides and assessments developed by states and local districts, widely accepted professional standards, scientific research, other types of research studies in the literature, key reports having significant national and international interest, international standards and assessments, other assessment instruments in the content area, and prior NAEP frameworks. (p. 7)

The universe of documents represented in this list is monumental for any given content area. No aspect of the process for selecting what to include in this library is specified. The NAEP framework policy provides some guidance on factors to "balance" in prioritizing source documents but is otherwise silent on the way that this library should shape panel deliberations and, ultimately, the framework being developed or reviewed.

The "Commissioning procedures" component of the *work process* element is not specified in any NAEP source reviewed.

As with *work product*, requirements for several aspects of the *work process* are specified in statements of work. Also, process reports of NAEP framework development or update [e.g., WestEd, 2006; WestEd, 2010; WestEd (draft), 2021] provide detailed schedules and accounts of meetings but only general statements about discussion topics, how consensus was reached, or how differences of opinion were addressed.

Owner, Timeframe, and Approval

The *owner* or client of NAEP assessment frameworks is the Governing Board. The *timeframe* for producing frameworks does not appear to be specified in general. Contract lengths or schedules in specific statements of work *report* desired timeframes.

The NAEP framework policy addresses the "Approving party" component of the *approval* element of framework processes. It does not specify an approval process or criteria for judging the quality of the *work process* or *product*. The policy does not specify the procedures to follow in case a framework project is not approved.

Pan-Canadian Assessment Program (PCAP)

The Pan-Canadian Assessment Program (PCAP) resembles NAEP in context: It is a national survey in a country without a single set of national-level academic standards or national curricula. The PCAP is given every three years in reading, mathematics, and science. PCAP was the first program that we reviewed, and this review greatly informed the development of our organizer for framework processes. Our review of this program is in Appendix B.

The SAT and the ACT

Two long-standing and well-recognized testing programs in the U.S. are the SAT and the ACT. Many colleges and universities require or accept these tests for admission. Recently, several states have adopted one or another of these tests to meet the ESEA requirement for testing in high school. The SAT is revised or redesigned every few years.

Due to these testing programs' national user base, the test takers they serve have been learning under different standards and curricula. Neither of these programs claims to be neutral with respect to curriculum, although the ACT more explicitly claims to incorporate information about the different curricula of the population of test-takers: Every three to five years, ACT conducts a national curriculum survey that asks K-12 and postsecondary educators to rate the importance of several discrete skills in their teaching or as a prerequisite to their course. ACT conducted the last such survey in 2020 (ACT, 2020 a).

Neither the SAT nor ACT programs provide detailed documentation of their assessment framework processes. ACT offers some highlights of the process in its most recent technical manual, particularly the sources or factors informing the ACT frameworks. These include subject-matter experts, academic research, ACT data, the ACT national curriculum survey, and a survey of other content standards – such as the Next Generation Science Standards (NGSS). (ACT, 2020 b, p. 1.6) However, most framework components listed in the organizer of this technical memo are not reported by ACT.

College Board documentation on framework processes for the redesigned SAT reveals a more hierarchical organization of committees and working groups involved in these processes. Their membership is not specified except in general terms (for example, "The Higher Education Advisory Working Group is composed of 30 representative higher education leaders from institutions across the nation." (College Board, 2015, p. 15). Available documentation on the input provided by these groups highlights *role* and not *process*. For example, "The group provides direct, in-depth feedback on such matters as implementation and reporting, scores and validation, and communications." (p. 15) Like the ACT, the SAT does not report on most framework process elements and their components.

Frameworks for State Assessments

Common Core State Standards (CCSS)

The Common Core State Standards (CCSS, NGA/CCSSO, 2010) are a seminal set of content standards in K-12 English language arts and mathematics, intentionally anchored in "college/career readiness," developed under the sponsorship of the National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO). Published in 2010, the CCSS were adopted by over 40 states, districts, and territories by 2013.

The CCSS are unusual in that their sponsorship by the NGA and CCSSO was as close to a set of "national, not federal" content standards created in modern times. The development process

involved four sets of contributors: a core team of lead authors that designed the architecture and key aspects of what became the CCSS, a "work team" heavily involved in writing the CCSS— first college/career readiness standards, and then K-12 standards— and several review groups, including an official "feedback group." There was also a "validation group" that considered the evidential and argumentative basis for the CCSS. And finally, multiple drafts of the CCSS were released for comment—both targeted (e.g., state departments of education, professional organizations) and public—and those comments were considered in creating the final versions of the CCSS. The lead authors and work groups for the CCSS were primarily university academics or people from business organizations; there was no specific call for active teachers or school administrators to be on the committees. None were, although some committee members had been elementary/secondary teachers previously, and several had worked with other sets of content standards. The "lead writers" consisted of three persons each for ELA and mathematics; the "work group" consisted of 24 total persons. The validation committee consisted of 29 members, primarily university- or institute-based academics, although there were also five teachers and principals, as well as a few employees of testing companies.

The CCSS were conceived as content standards for instruction, not assessment specifications. The intent of the CCSS—for example, for assessment—was commented on by individual lead authors and by an organization established by a few of the CCSS lead authors—Student Achievement Partners. However, these were not treated as authoritatively reflecting the consensus of the CCSS authors and development process. States and others developing assessments were able to treat the CCSS as academic content standards and develop different assessment constructs, blueprints, and other specifications. For example, two federally funded consortia, each joined by many states, developed quite different assessment specifications using quite different development processes, resulting in the two different operational assessments by the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium.

There is little documentation available regarding the processes of how the committee number, structure, or membership were determined; or the processes by which the CCSS were conceptualized or developed in terms of how committee work was allocated, how leadership took place, or differences reconciled. Also, although a public comment process was engaged in by the developers of the CCSS, we could not find documentation of the process by which comments were solicited or responded to. Some of this may be attributed to the fact that NGA, CCSSO, and the work groups wanted to control the development without undue outside influence until formal feedback was instituted. Some may also be attributed to the subsequent controversial nature of the CCSS; for example, neither NGA, CCSSO, nor the website they established for the Common Core have listings of the various committee members, let alone primary documentation of the CCSS developmental process on their websites.

Next Generation Science Standards (NGSS)

The Next Generation Science Standards (NGSS) are a widely popular set of K-12 science content/assessment standards. Over 30 states had adopted some version of the NGSS by 2021. The NGSS have two foundational documents: A framework document and a standards document, authored and published independently.

The *Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (National Research Council, 2012) was authored by a group sponsored by the National

Research Council, National Academy of Sciences.[3] The committee responsible for the *Framework* consisted of 18 persons, including "practicing scientists, including two Nobel laureates, cognitive scientists, science education researchers, and science education standards and policy experts." (Achieve, n.d. a) There was no charge for specific groups to be represented on the writing committee; no elementary/secondary educators were included.

The Framework document included applications of the *Framework* to specific science domains. "In addition, the NRC used four design teams to develop the *Framework*. These four design teams, in physical science, life science, earth/space science, and engineering, developed the *Framework* sections for their respective disciplinary area." (Achieve, n.d. a) The development process included gathering public comments. "After releasing a public draft in July of 2010, the NRC reviewed comments and considered all feedback prior to releasing the final *Framework*." (Achieve, n.d. a)

The *Next Generation Science Standards* document provides specific content standards reflective of the *Framework* in grades K-5, middle school, and high school. Thirteen appendices provide additional information regarding rationale, additional information, and discussion of relevant issues in extending the Framework into Standards. The Standards were produced by a group of 26 Lead State Partners, managed by Achieve (Achieve, n.d. b). One of the key Achieve staff persons and another member of the NGSS writing team had been members of the *Framework* committee. The writing committee for the Standards included many state department of education employees, but there was not a charge for specific representation from specific groups. Educator input was specifically and actively sought during the feedback and comment processes.

The NGSS have a conceptual *Framework* document developed separately from the *Standards* document. One disadvantage is that the two committees were not together to work out issues. A prime example is that the *Framework* delineates a domain much larger than is possible to assess practically, or even perhaps to learn. The developers of the *Standards* had to make choices about what to include and what to leave out, without the authoritative agreement of the *Framework* authors. And although the authors of the *Standards* aimed them at assessment specifications, they worked at the level of individual standards rather than defining what would be adequate for a construct or domain. The result has been that states that have adopted the NGSS have adopted different things: notably, some have adopted the *Framework*, while other have adopted the *Standards*; some consider the performance expectations in the *Standards* to be the standards, while others consider the performance expectations merely examples. States and their partners have struggled to use the documentation to create practical assessment blueprints, and there has been considerable variation across states.

The NGSS publicly available documentation does not include information regarding the processes of how the committee number, structure, or membership was determined; or the processes by which the *Framework* or *Standards* were conceptualized or developed in terms of how committee work was allocated, how leadership took place, or differences reconciled. Also, although a public comment process was engaged in by the developers of both the *Framework* and the *Standards*, documentation did not include detailed description of the process by which comments were solicited or responded to.

[3] A starting point for documentation about the Framework development is https://www.nationalacademies.org/our-work/conceptual-framework-for-new-science-education-standards#sectionCommittee

International Assessments

The assessment frameworks of the two leading international assessment programs have very different conceptual relationships to curricula.

Programme for International Student Assessment (PISA)

The Programme for International Student Assessment (PISA) is a sample-based assessment headed by the Organization for Economic Cooperation and Development (OECD) and administered to 15-year-olds in participating countries and economies (79 in 2018) once every three years. The first PISA assessment was in 2000. Domains assessed include reading, mathematics, science, and financial literacy. PISA assesses an innovative domain in each cycle. In 2018, that was global competence (OECD, 2019). PISA does not purport to align to any curricular or content standards. Instead, it aims to assess "the extent to which 15-year-old students near the end of their compulsory education have acquired the knowledge and skills that are essential for full participation in modern societies." (OECD, 2019, p. 11). The PISA Governing Board (OECD, n.d.) has members from each participating country. Framework and related documents are available through the PISA website.

The most recently published framework (for 2018, when reading was the "major domain" assessed) lists the chair and members (total of 6) of the reading framework working group. The same information is provided for the global competence working group (total of 5). All members are affiliated with universities or similar organizations. The global competence framework was developed by a member of the OECD Secretariat working with a university collaborator (OECD, 2019, pp. 18-19). Publicly available documents do not indicate which, if any, elements or components of framework processes were shaped by requirements specifications. The *work process* components are not reported.

Trends in International Mathematics and Science Study (TIMSS)

The Trends in International Mathematics and Science Study (TIMSS) has been assessing mathematics and science in fourth and eighth grade every four years since 1995. In 2019 – the most recent year of administration – 64 countries and 8 "benchmarking participants" (generally, cities) participated in TIMSS (Mullis et al., 2020). TIMSS assesses mathematics and science in grades 4 and 8.

The TIMSS assessment frameworks highlight the importance of curriculum as the basis for the domain description. The most recent assessment frameworks indicate they are updates of earlier frameworks. Framework documents list names of members of the framework revision committees. These also serve as members of item review committees. In the most recent revision of the TIMSS framework (2019), there were 7 members per content area; most are university staff and are described as "internationally recognized mathematics and science experts." (Mullis & Martin, 2017, p. 96). However, the frameworks also present an extensive list of TIMSS national research coordinators (at least one per participating country) who "participated in a series of reviews of the updated frameworks." (p. 98) As with PISA, available documents (assessment frameworks, technical reports, etc.) do not indicate which, if any, elements or components of framework processes were shaped by requirements specifications. The *work process* components are not reported.

Professional Standards and Framework Processes

Processes for framework development are not covered extensively in widely available professional standards that deal with test development or validation. The *Standards for educational and psychological testing* (*Standards*, AERA/APA/NCME, 2014) address select aspects of framework processes in Chapter 4, Test design and development, Test specifications (pp. 75-81). In the *Standards*, test development begins with developing test specifications. In many ways, this places the framework processes beyond the scope of the *Standards* because the essential component of assessment frameworks (the domain description) precedes test specifications. Note, however, that most assessment frameworks contain at least some assessment design aspects. The *Standards* apply to *these* parts of assessment frameworks and thus framework processes more generally:

> The term *test specifications* is sometimes limited to description of the content and format of the tests. In the *Standards*, test specifications are defined more broadly to also include documentation of the purpose and intended uses of the test, as well as detailed decisions about content, format, test length, psychometric characteristics of the items and test, delivery mode, administration, scoring, and score reporting. (p. 76)

The Standards have little to say about appropriate processes for deriving domain descriptions (also called *content specifications* and *content frameworks* in the *Standards*) for achievement tests such as NAEP: “The delineation of the content specifications can be guided by theory or by an analysis of the content domain (e.g., an analysis of job requirements in the case of many credentialing and employment tests).” (p. 76)

The *ETS Standards for quality and fairness* (ETS, 2015) closely follow the *Standards* and do not explicitly address framework processes. One ETS standard speaks to settings where information about the construct is not readily available, indicating that “obtaining the information may be part of the test developers’ (typically, a contractor) task.” The standard continues, “If the information has to be obtained, work collaboratively with clients, subject-matter experts, and others as appropriate.” (p. 29) But the ETS standards go no further in discussing appropriate framework processes.

Guidance published by the Department of Education for the assessment peer review process addresses some requirements for state (Every Student Succeeds Act, ESSA) assessment framework processes. State assessment programs must show that they have “challenging academic content standards in reading/language arts, mathematics, and science” that are “aligned with entrance requirements for credit-bearing coursework in the system of public higher education in the State and relevant State career and technical education standards.” (U.S. Department of Education, 2018, pp. 30). Among the examples evidence that states can provide to meet this requirement, the guidelines cite:

> A detailed description of the strategies the State used to ensure that its academic content standards adequately specify what students should know and be able to do;
>
> Documentation of the process used by the State to benchmark its academic content standards to nationally or internationally recognized academic content standards; Reports of external independent reviews of the State’s academic content standards by content experts, summaries of reviews by educators in the State, or other documentation to confirm that the State’s academic content standards adequately specify what students should know and be able to do;

> Endorsements or certifications by the State's network of institutions of higher education (IHEs), professional associations and/or the business community that the State's academic content standards represent the knowledge and skills in the content area(s) under review necessary for students to succeed in college and the workforce. (pp. 30-31)

These examples suggest some principles or standards for framework processes in the context of ESSA, especially around vetting or approval. However, this is a special context in which there is an independent criterion (college and career readiness) built into the mandate for ESSA.In either case, there is a principle implied by the peer review guidance: When there is an external referent in the mandate, then framework development should incorporate some process to ensure that the content to be assessed is related to that criterion.

The previously referenced NAEP framework policy (Governing Board, 2018) comes closer to supplying professional standards for framework processes than any other source. Principles 1 (Elements of Frameworks) and 5 (Elements of Specifications) address some of the components of the framework process element *work product*. Similarly, some components of *work process* are addressed in Principles 2 (Development and Update Process), 3 (Framework Review), and 4 (Resources for the Process). Principle 6 (Role of the Governing Board) covers components of *work process*, *owner*, and *approval*.

Key Findings

Five elements of framework processes answer foundational questions about framework development. These elements are: The conditions for initiating a framework (or review), what is to be included in a framework, what are the steps or rules to be followed in putting a framework together, who owns the framework process, what is the timeline for the process, and what is the process for approval.

There is considerable variation among assessment programs in the framework process elements that programs report. Some programs specify general requirements for some elements (or components thereof). No program we know of specifies requirements for all components.

Although most programs have a structure for framework development, such as a sequence of panels or working groups, no assessment program we reviewed specifies systematic processes for (a) selecting panel members or authors, (b) selecting source documents, (c) addressing competing views about what should be in the framework, (d) integrating source documents, expert judgment, and public review to derive a framework, and (e) approving the final product, together with a contingency plan in case the work is not approved.

Implications of NAEP Legislative Mandate for NAEP Framework Processes

Here we address implications of three aspects of NAEP law and tradition: Curricular neutrality, representation of diverse views, and the role of professional standards.

Curricular Neutrality

By tradition and by law, NAEP has been guided by a criterion of curricular neutrality.

The concept is applied to framework processes in NAEP's framework development policy statement, which includes as a guideline that:

> The framework shall focus on important, measurable indicators of student achievement to inform the nation about what students know and are able to do without endorsing or advocating a particular instructional approach. (Governing Board, 2018, p. 4)

However, the standards, curriculum, and teaching practices in the U.S. are relevant to the NAEP framework, even if NAEP adopts a neutral stance. (See, for example, the list of resources that the NAEP framework policy Principle 4 asks panelists to consider.)

The principle of curricular neutrality has implications for the NAEP framework development process. Whatever those may be, they are not explicit in the NAEP documentation we reviewed. Among our recommendations for future work, we offer some considerations towards more precise definition of curricular neutrality to inform framework processes on a NAEP-wide level.

Diversity of Views

The NAEP framework policy indicates that framework panels "shall reflect diversity in terms of gender, race/ethnicity, region of the country, and viewpoints regarding the content of the assessment under development." (Governing Board, 2018, p. 5)

Ensuring representation of diverse viewpoints regarding assessment content implies that the process for selecting framework panel members should be informed of both existing viewpoints and candidate panelists' views. It may be that in practice, this is or has been part of the panelist selection process.

"[D]iversity in terms of […] viewpoints regarding the content of the assessment" would likely include experts who have strong opinions not only about the nature of the construct but also about the appropriateness, for their content domain, of measures largely composed of multiple-choice test items.

The representation of diverse viewpoints on panels is likely to result in perspectives that cannot always be reconciled into one framework. How should impasses be handled? Rules of order might be specified ahead of time.

Role of Professional Standards

NAEP law references "professional standards" or "professional assessment standards" several times. Three instances have implications for framework processes. In the first, "professional standards" are referenced as the basis for the development of "assessment objectives," "test specifications," or both:

> IN GENERAL – In carrying out its functions under this section the Assessment Board shall—[…] develop assessment objectives consistent with the requirements of this section and test specifications that produce an assessment that is valid and reliable, and are based on relevant widely accepted *professional standards* [Section 302, (e)(1)(C), emphasis ours]

The second and third instances concern the determination of achievement levels:

> IN GENERAL- Such levels shall-- be determined by—(I) identifying the knowledge that can be measured and verified objectively using widely accepted *professional assessment standards*; and (II) developing achievement levels that are consistent with

> relevant widely accepted *professional assessment standards* and based on the appropriate level of subject matter knowledge for grade levels to be assessed, or the age of the students, as the case may be. [Section 303, (e)(2)(A)(i)(I-II), emphasis ours]

The importance of professional standards is evident in the NAEP law. However, a central question is to what extent do they apply to framework processes as understood in this technical memo? If they apply at all, then the lack of a robust set of professional standards for framework processes poses a real challenge for assessing the extent to which any NAEP program involving framework processes was properly designed and implemented.

How this Review Might Inform NAEP Framework Processes

This review might inform NAEP framework processes primarily through the organizer we developed. We believe that all elements and components should certainly be documented for any framework project. More importantly, the NAEP program may benefit from more deliberate consideration of the extent to which it wishes to specify requirements for those components, and whether (or when) it will delegate such requirements specification to others, such as contractors.

Delegation of requirements specification may lead to different requirements for different testing programs. This may be appropriate for some elements/components – for example, insisting on content-by-process organization of all domain descriptions could run counter to current or future conceptualizations of domains. But there doesn't seem to be an obvious rationale for diverse requirements specifications for some other components, such as all *work process* components.

Towards Best Practices for Framework Processes

The absence of professional standards for most components of framework processes leaves much room for proposing principles, guidelines, and standards.

We propose that sponsors make deliberate choices regarding which components to specify requirements for and to document the rationale for those choices.

When sponsors consider delegating requirements specification for a component to other groups or contractors, it may be useful to prepare for the different ways in which the component may unfold, possibly resulting in very different work products.

A good analogy for what a systematic framework development process might look like is standard-setting. There are many standard-setting methods, and no consensus about which is best in every case. However, the more mature methods prescribe a step-by-step process, contingency planning, specific documentation requirements, and success criteria. Disagreements are addressed through rounds of conversation and voting procedures.

As with standard-setting, it may be possible to outline a standard set of procedures for some special cases of framework development.

Standard-setting needs an external criterion, or has to very heavily rely on process and internal coherence. A reliance on what has sometimes been called "procedural validity"—that is, the quality and evaluation of quality are dependent upon having a good process—needs to show reasonable process for producing work products and evaluation showing implementation fidelity.

For example, suppose that (by sponsor-level specification or by contractor-level specification) it is decided that the process for generating NAEP assessment objectives will involve sub-setting from a broader set of content standards. One can imagine a few ways to approach this general task, involving discussions and voting. Those approaches can be cast as systematic framework development methods.

When the sources are many and varied and the actual task of creating a framework less certain, sponsors can still indicate how each type of source should inform framework development. Sponsors might also specify what the resulting assessment objectives should look like individually – in terms of syntax, length, the extent of performance description (see "content/performance continuum" in the section on recommendations for additional work), and similar properties – as well as collectively.

Recommendations for Additional Work to Inform Governing Board Considerations

This section proposes additional studies, reviews, or conceptual work to help inform how the Governing Board addresses framework processes. We elaborate on some of the proposals.

Proposal 1. Every assessment program has a definition or description of the domain to be assessed; this is part of every assessment framework. (See framework process element *work product*, component "Domain description.") There is considerable variation in how frameworks arrive at these descriptions, however. The Governing Board might explore the structure of domain descriptions in different assessment frameworks to decide which is most appropriate NAEP-wide.

Proposal 2. Review the different kinds of sources informing assessment frameworks to develop a systematic way to incorporate those sources into the framework development process.

Commentary. One class of sources includes content standards that may differ in terms of their educational orientation.

All assessment frameworks report domain descriptions that are assessment-oriented. This means that they were developed for the purpose of creating an instrument to determine what students know and can do. By contrast, domain descriptions can be oriented toward instruction – that is, primarily for the purpose of getting students to know and be able to do the knowledge/skills that are indicated. Some content standards, such as the high-level academic content standards that states adopt, purport to inform both uses. The sources from which an assessment framework might draw may be instruction-oriented, assessment-oriented, over-arching, or some combination of these.

Academic content standards adopted by states are good examples of over-arching domain descriptions: States typically adopt content standards to specify what, at a minimum, students should learn and be able to do. These content standards are intended to provide guidance for educators as they select or develop curricula and as they design their associated instruction. Instructional and over-arching domain descriptions generally encompass more than those for large-scale assessments.

Domain descriptions for instruction include more than those for assessment in that the former often specify:

- More complex content than can feasibly be assessed in large-scale assessments – such as the full writing process, including research projects; and
- Skills that do not fit well within the tradition of assessment of work products produced by individuals working alone, such as mental math, problems solved in groups, cross-curricular learning targets, non-standardized learning targets such as individual projects, and learning arising from extended experiences such as reading specific novels in a literature class.

The content standards that go into a domain description for assessment will typically be a subset of over-arching standards or those with a (primarily) instructional orientation.

Whenever the process for generating a domain description in an assessment framework involves sub-setting from a broader set of content standards for learning, the sponsors for an assessment program might specify how that is done (element *work process*, component *sources*). At minimum, they should require that the process by which it is done be documented (element *work process*, component *documentation requirements*). For transparency purposes, the sponsor may require that this documentation be included in the framework itself (element *work product*, component *documentation of process*).

Proposal 3. Consider the *content/performance continuum* of assessment objectives, to specify which is most appropriate for NAEP.

Commentary. In most assessment programs, the foundational unit of content specifications (typically found in assessment design documents) is called a "content standard." However, there is considerable variation in what is included in a content standard across assessment programs. Content standards always contain the content of the construct (if the construct is a skill, the description of that skill to be assessed would be the "content" of the content standard). Important variations occur around what else is included in the content standard—particularly, how much of a performance description is included in the content standard.

Content standards used by assessment programs can be classified on a continuum reflecting increasingly elaborate performance descriptions. Assessment sponsors can choose to specify in advance where on this continuum to target the resulting content standards, and direct assessment framework authors to write frameworks in such a way that assessment content standards derived from those frameworks will be at their chosen level:

1. Content only. The content standard describes what students should know or understand or be able to do but does not include how a student is supposed to demonstrate that knowledge, understanding, or skill.

2. Content with minimal performance descriptions. The content standard includes description of the content and indicates what the student is supposed to be able to do with that knowledge, understanding, or skill. Minimal detail is provided in this performance description. Very many U.S. state content standards use this structure.

3. Content with detailed performance descriptions. The content standard includes description of the content and indicates in some detail what the student is supposed to be able to do with it or how the student is supposed to demonstrate the desired level of expertise. The Next Generation Science Standard's (NGSS) Performance Expectations (P.E.s) are a widely known example of this approach.

4. Content with multiple detailed performance descriptions at different levels. The content standard includes content and descriptions of multiple levels of expertise and/or how the student demonstrates those levels of expertise. Examples of content standards using this approach include those developed in the "learning progressions" approach. Dynamic Learning Maps (DLM) precursors and NWEA for Nebraska range ALDs employ this approach.

This aspect of the structure of content standards has far-reaching implications for assessment specifications, designs, and activities. NAEP can choose to specify what to include about it, both in terms of content and process, in its framework process guidance across programs. This would lead to assessment content standards written at parallel levels of specificity across content areas.

Proposal 4. Explore the ways in which assessment programs attempt to remain "neutral" with respect to curriculum, to state how NAEP will provide guidance (requirements specification) so its resulting assessment frameworks are all "curriculum neutral" in the same ways.

Commentary. Most large-scale U.S. state assessments aim to be more general than a specific curriculum. States resolve this issue through the mechanism of common content standards. Other contexts, such as some national and all international assessment programs, however, operate across jurisdictions with different curricular/content standards. These programs also aim to be more general than a specific set of curricular/content standards, and thus must adopt some conceptual relationship to the curricula/content standards of the assessed population.

How they go about that varies. Some programs, such as PCAP, provide a general criterion (what is common across the curricula for the different jurisdictions in the population tested). However, PCAP does not go further in specifying how that commonality is to be judged or determined. NAEP does not provide a specific criterion, nor a specific process for considering the curricula (or academic content standards) of the assessed population.

Some approaches to help ensure an assessment is not tied too closely with a particular curriculum or state content standards:

- Determine what is common across the curricula/content standards of the assessed population. An assessment may focus on those things which all curricula agree on; that might be found through a systematic survey of relevant curricula. This is done explicitly for at least one non-U.S. assessment program. (We note that NAEP also has conducted such studies but, to our knowledge, not expressly to test what is common.) Note that the methodology for determining what is common, and assessing whether the process results in something meaningful, is a separate and non-trivial matter that could be addressed ahead of time.

- Refer to education research in the content domain and deliberately ignore curricula/content standards. An assessment may build its content specifications from research only, if available, without referencing curricula. If the research literature is extensive and detailed enough, it may provide sufficient basis to generate content standards, especially if there is broad consensus about the research base. Note: This seems like the least practical to us and the most difficult to specify requirements for. We include it here anyway for completeness.

- Refer to other authoritative content frameworks, without referencing curricula. If there is a widely accepted content framework outside the assessment program, that content framework may be adopted for the assessment program, especially if that content framework does not reference specific curricula. This is what was done by states adopting the Common Core State Standards, the Next Generation Science Standards, and other content standards generated by national or professional consensus such as the NCTM content standards and the previous National Science Standards. There is at least one challenge for NAEP here: An assessment framework derived from an authoritative content framework is difficult to distinguish from an assessment framework for the curriculum implied by that authoritative content framework (and thus potentially not "curriculum neutral").
- Refer to international assessment frameworks for assessments in which many countries participate. Some challenges: (1) How would NAEP not simply be a different instantiation of that international program? And is it a problem if it were? (2) This option may or may not be consistent with different readings of the NAEP law. (3) There are likely strong political views, pro and con, about the relevance of education in other countries to an assessment of educational progress for U.S. students. What is the scope of NAEP's curricular relevance/neutrality? Is it curricula in the U.S. or curricula throughout the world?

Proposal 5. Study what goes into the assessment design component of frameworks for different assessment programs and consider whether developing test specifications should also be part of the framework development task involving the same group or groups.

Commentary. There typically are two levels of specifications for assessments. One level is more foundational. The other is more detailed. The more foundational may be thought of as defining the core validity claims for the assessment, while the other level specifies how those claims are to be supported in terms of assessment evidence. In many large-scale assessment programs, such as state assessment programs, there is an explicit division in who is responsible for developing which level of specifications. The state is explicitly responsible for developing the first level of specification without input from possible vendors, because the first level of specifications often constitutes the core of a request for proposals. Bidders then propose the second set of specifications—or how to develop them—as the vendor's responsibility. Of course, the vendor's proposals must be approved by the program sponsor; often there is iterative consultation between the program sponsor and vendor to arrive at this second level of specification. Explicit in this organization is the assumption that there are multiple possible ways the second level can be specified, once work at the foundational level is complete. Some of those ways may not reflect the intentions of those who developed the foundational level frameworks.

Proposal 6: Investigate best practices for including implementation fidelity evaluation and documentation.

Commentary. Since NAEP's development of assessment frameworks are so dependent on processes being specified and followed well, the development process might benefit from incorporating means to formatively check on the quality of the process while the framework is being developed, as well as a summative evaluation. For example, if the purpose of recruiting a diverse committee is to ensure diverse perspectives contribute to the framework development, then a formative evaluation would check whether committee members feel comfortable during the process. This could be accomplished through a survey with items such as, "I feel my voice is

being heard," "I am clear about the objectives of our committee work," "The work is well-organized," "I think committee assignments are fair," etc. An external evaluator could support the formative evaluations. Similarly, a summative evaluation should include evaluation of the process. This should incorporate documentation of "procedural validity" that would support the quality of the assessment framework. The summative evaluation of the process should also draw lessons learned to help inform future NAEP assessment frameworks.

Proposal 7: Draw on the best available knowledge to inform effective committee work, especially processes for generating, discussing, and resolving issues.

Commentary. A review of the research literature and professional practice should be able to inform different ways to deal with power dynamics—how to ensure all contribute as intended by inclusion in representation, such as how to structure discussions, when to use open versus anonymous voting, etc. There may be different group dynamics and methods to produce a group report when there is more or less agreement about fundamental issues. It would have to be decided how best to make such information available to the committees.

References

Asterisked documents are in the set provided by the Governing Board for this review.

Achieve. (n.d.). Next Generation Science Standards: Development Overview. Webpage: https://www.nextgenscience.org/development-overview

Achieve. (n.d.). Next Generation Science Standards. Developing the Standards. Webpage: https://www.nextgenscience.org/developing-standards/developing-standards

ACT. (2020 a). ACT National curriculum survey 2020. Available: http://www.act.org/content/dam/act/unsecured/documents/NCS_Report_Web_2020.pdf

ACT. (2020 b). ACT Technical manual. Version 2020.1. Available: https://www.act.org/content/dam/act/unsecured/documents/ACT_Technical_Manual.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

College Board. (2015). Test specifications for the redesigned SAT. Available: https://collegereadiness.collegeboard.org/pdf/test-specifications-redesigned-sat-1.pdf

ETS. (2015). *ETS Standards for quality and fairness*. Available: https://www.ets.org/s/about/pdf/standards.pdf

*Martineau, J.; Dadey, N., & Marion, S. (2018). *Literature review on developing or revising assessment frameworks to support a transition from paper-based to digitally based assessment*.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/

*National Assessment of Educational Progress Authorization Act of 2002. P.L. 107-279. Available: https://www.nagb.gov/about-naep/the-naep-law.html

National Assessment Governing Board. (November 17, 2018). National Assessment Governing Board Developing student achievement levels for the National Assessment of Educational Progress policy statement. Available: https://www.nagb.gov/content/dam/nagb/en/documents/policies/ALS-revised-policy-statement-11-17-18.pdf

*National Assessment Governing Board. (March 3, 2018). National Assessment Governing Board Framework development policy statement. Available: https://www.nagb.gov/content/dam/nagb/en/documents/policies/framework-development.pdf

*National Assessment Governing Board. (May 11, 2018). Update of the National Assessment of Educational Progress (NAEP) frameworks for mathematics, reading and other subjects. RFP# 91995918R0002. Attachment A: Statement of work.

*National Assessment Governing Board. (n.d.). Summary chart of framework updates.

National Governors Association Center for Best Practices, Council of Chief State School Officers. 2010. Common Core State Standards. National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington D.C.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Organization for Economic Cooperation and Development. (2019). "What is PISA?", in *PISA 2018 Assessment and Analytical Framework*, OECD Publishing, Paris. Available: https://doi.org/10.1787/2c7c311d-en

Organization for Economic Cooperation and Development. (n.d.). PISA Governing Board. Webpage: http://www.oecd.org/pisa/contacts/pisagoverningboard.htm

U.S. Department of Education. (July 2018). A state's guide to the U.S. Department of Education's assessment peer review process. Available: https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf

*WestEd. (May 19, 2006). Final report: NAEP Science assessment 2009.

*WestEd. (December 6, 2010). Final report: Technology and engineering literacy assessment for the 2014 National Assessment of Educational Progress.

*WestEd. (December 6, 2019). Design document: 2025 NAEP mathematics framework update.

WestEd. (draft). (February 12, 2021). Process report: 2025 NAEP Mathematics framework update.

Appendix A: Additional Documents Reviewed

Asterisked documents are in the set provided by Governing Board for this review. Double-asterisked documents are those consulted during the PCAP review.

*Boyer, M. (August 2020). Enhancing the meaning and utility of NAEP results through linking with other assessments.

**Council of Ministers of Education, Canada (CMEC). (1997). *Common Framework of Science Learning Outcomes K to 12*. Available: https://archive.org/details/commonframework00coun

**Council of Ministers of Education, Canada (CMEC). (2008). PCAP-13 2007: Report on the Assessment of 13-Year-Olds in Reading, Mathematics, and Science. (T.Y. 2007). Retrieved from: https://www.cmec.ca/Publications/Lists/Publications/Attachments/124/PCAP2007-Report.en.pdf

**Council of Ministers of Education, Canada (CMEC). (2011). PCAP-2010 Report on the Pan-Canadian Assessment of Mathematics, Science, and Reading. (T.Y. 2010). Retrieved from: https://www.cmec.ca/Publications/Lists/Publications/Attachments/274/PCAP2010.pdf

**Council of Ministers of Education, Canada (CMEC). (2014). PCAP 2013 Report on the Pan-Canadian Assessment of Science, Reading, and Mathematics. (T.Y. 2013). Retrieved from: http://cmec.ca/Publications/Lists/Publications/Attachments/337/PCAP-2013-Public-Report-EN.pdf

**Council of Ministers of Education, Canada (CMEC). (2015). Agreed memorandum on a council of ministers of education, Canada. Retrieved from: https://www.cmec.ca/Publications/Lists/Publications/Attachments/360/2016-08-17-Agreed-Memorandum-EN.pdf

**Council of Ministers of Education, Canada (CMEC). (2016). Pan-Canadian assessment program 2016 assessment framework. Retrieved from: https://www.cmec.ca/docs/pcap/pcap2016/PCAP-2016-Assessment-Framework.pdf

**Council of Ministers of Education, Canada (CMEC). (2018). PCAP 2016 Report on the Pan-Canadian Assessment of Reading, Mathematics, and Science. (T.Y. 2016). Retrieved from: https://www.cmec.ca/Publications/Lists/Publications/Attachments/381/PCAP-2016-Public-Report-EN.pdf

**Council of Ministers of Education, Canada (CMEC). (2019). PCAP 2016 Technical Report. (T.Y. 2016). Retrieved from: https://www.cmec.ca/Publications/Lists/Publications/Attachments/394/PCAP%202016%20Technical%20Report_FINAL_EN.pdf

**Council of Ministers of Education, Canada (CMEC). (2020). Pan-Canadian assessment program 2019 assessment framework. Retrieved from: https://www.cmec.ca/docs/pcap/pcap2019/PCAP-2019-Assessment-Framework-EN.pdf

**Council of Ministers of Education, Canada (CMEC). (n.d. a). Pan-Canadian assessment program (PCAP) mathematics assessment framework. (T.Y. 2010). Retrieved from: https://www.cmec.ca/docs/pcap/pcap2013/Math-Framework-April-2013-EN.pdf

**Council of Ministers of Education, Canada (CMEC). (n.d. b). Pan-Canadian assessment program (PCAP) reading assessment framework. (T.Y. 2007). Retrieved from: https://www.cmec.ca/docs/pcap/pcap2007/Reading-Framework-EN-Apri2013.pdf

**Council of Ministers of Education, Canada (CMEC). (n.d. c). Pan-Canadian assessment program (PCAP) science assessment framework. (T.Y. 2013). Retrieved from: https://www.cmec.ca/docs/pcap/pcap2013/PCAP-2013-Science-Assessment-Framework-EN.pdf

**Council of Ministers of Education, Canada (CMEC). (n.d. d). Pan-Canadian assessment program website: https://www.cmec.ca/240/Pan-Canadian_Assessment_Program_(PCAP).html

*Dickinson, E.R., Gribben, M., Schultz, S.R., Spratto, E., & Woods, A. (February 2, 2020). Comparative analysis of the NAEP science framework and state science standards: Technical report.

*Jago, C. (March 2009). A history of NAEP assessment frameworks.

*Johnston, W.T., Stephens, M., & Ratway, B. (September 2018). Review of state curricular standards in mathematics: Supplemental exhibits.

**Kelly, D.L., Centurino, V.A.S., Martin, M.O., & Mullis, I.V.S. (Eds.) (2020). *TIMSS 2019 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/encyclopedia/

**Rostamanian, A. (2020). Canada. In Kelly, D.L., Centurino, V.A.S., Martin, M.O., & Mullis, I.V.S. (Eds.), *TIMSS 2019 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/encyclopedia/

*Stephens, M., Johnston, W.T., & Ratway, B. (December 2018). Review of state curricular standards in mathematics: Methodological narrative.

**Wendt, H., Schulz-Heidorf, K., Walter, D., & Steffensky, M. (2020). Germany. In Kelly, D.L., Centurino, V.A.S., Martin, M.O., & Mullis, I.V.S. (Eds.), *TIMSS 2019 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/encyclopedia/

Appendix B:
Review of Framework Processes in the Pan-Canadian Assessment Program (PCAP)

Relevance of PCAP

According to the *TIMSS 2019 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*, the U.S. is not the only participating country without a national mathematics or science curriculum. Other countries without national curricula in these subjects in grade 4 include Belgium (Flemish), Bosnia and Herzegovina, Canada, and Germany (Kelly et al., 2020, Introduction p. 7). Among these four countries, only Germany has national education standards that are binding across the primary divisions of the country. In general, each of Germany's 16 federal states, however, has a different curriculum aligned to those standards (Wendt et al., 2020, Germany p. 1).

In this list of countries without national curricula, only the U.S. and Canada have a national assessment, and in Canada, it is only at grade 8. This assessment, known as the Pan-Canadian Assessment Program (PCAP), assesses student achievement in reading, mathematics, and science. Like NAEP, participation in PCAP is based on random sample selection (Rostamanian, 2020, Canada p. 8).

Assessment Frameworks

The Council of Ministers of Education, Canada (CMEC) oversees PCAP. Documentation on this assessment program is available on the CMEC website (CMEC, n.d. d). The first administration of PCAP was in 2007, following a CMEC directive that "a new pan-Canadian assessment program was needed to reflect changes in curriculum, integrate the increased jurisdictional emphasis on international assessments, and allow for the testing of the core subjects of mathematics, reading, and science." (CMEC, n.d. d). PCAP has been administered every third year since 2007.

CMEC provides a PCAP assessment framework document for each of these administrations. These documents each describe one or more of four frameworks in the PCAP programs (reading, mathematics, science, and questionnaire). In the most recent assessment framework published (for 2019, CMEC, 2020), there is a chapter dedicated to each of the four frameworks. Each of these chapters includes a description of its subject framework, variously characterized as a "working definition" (mathematics), "definition" and "organization of the domain" (science), "definition" following a "theoretical background" (reading), and "description" followed by "core questions" (questionnaire).

The 2019 PCAP framework document has a 6-page introduction to the PCAP, its contrast with classroom assessments, its languages and modes of administration, reporting aspects, and monitoring role. The document closes with a 3-page chapter on assessment design, briefly covering scale characteristics, administration time, numbers of booklets, descriptions of item types (selected response and constructed response), and item release schedules.

The framework document from the 2016 cycle of PCAP contains much of the same information. Although PCAP assessed students on all three subjects starting in 2007, the frameworks for a given content area do not appear prior to the year it was first a "primary" domain for PCAP (2007 for reading, 2010 for mathematics, and 2013 for science). The framework documents for those years, moreover, cover only the framework of the "primary" domain. Thus, the text for the reading framework first appears in 2007, then again, with some updates and variations in the

2016 assessment frameworks document and again (with some changes) in the document for 2019.

Key Aspects of Framework Processes

These documents, together with information on the PCAP section of the CMEC website, as well as public and technical reports published through the 2016 cycle (except for 2007, which does not have a technical report), are collectively called the "program documentation" here. Program documentation describes some of the processes for developing the PCAP frameworks. They leave some aspects of framework processes unaddressed.

Authority and/or Legislative Mandate

There is no legislative mandate for the administration of PCAP. Authority over the program is exercised by the CMEC, whose members are the provincial/territorial education ministers of Canada. CMEC is governed by a memorandum; this agreement does not explicitly address standards, curriculum, instruction, or assessments among its objectives or duties. The CMEC memorandum, however, lists that the Council "may conduct and support research and cross-jurisdictional assessments." (CMEC, 2015, p. 2)

There is no readily available official agreement currently governing the PCAP program. The first PCAP public report (CMEC, 2008) indicates that CMEC convened an August 2003 PCAP working group which commissioned a "concept paper […] that would elaborate on issues of structure, development planning, operations, and reporting" (p. 2) The report does not cite this concept paper. The report states, however, that the working group used it to define the PCAP, a definition followed by six brief bulleted statements addressing (among other aspects) assessed domains, population, frequency, basis ("the commonality of all current juristictional [sic] curricular outcomes across Canada", p.2).

Descriptions of Framework Derivation Process

None of the PCAP sources offer a description of how a person or group derived the current frameworks .

Intended Relationship to Academic Standards or Curricula of the Assessed Population

Sources indicate that the PCAP frameworks are informed by the curricular goals/objectives/outcomes of the participating provinces/territories. Each content area framework and public report either states or implies that the PCAP frameworks cover what is common across participants' curricular goals/objectives/outcomes.

Role of Curricula/Content Standards of the Assessed Population

Each content area framework indicates it is informed by one or two of three kinds of external sources. The first kind, addressed by all three frameworks, concerns the curricula of the participating provinces/territories. The mathematics and science frameworks each reference reviews, authored by CMEC and not published, comparing the curricula of that content area, across Canada. The reading framework implies that a review was conducted, but only refers the reader to official jurisdictional websites for updated curricula.

Role of Education Research in the Content Area

The second kind of external source concerns education research in the content area. For the reading framework, it is "current research findings and best practices in the field of literacy development and the learning of reading." (n.d. b, p. 1). The original reading framework (from the cycle 2007 assessment) does not cite one specific document that summarizes the relevant education research, but instead provides the author's (or authors') own view(s) about the domain of reading, citing several other sources, primarily in reading/literacy theory. The domain description section of the reading framework chapter of the cycle 2016 assessment framework document (CMEC, 2016) is a significantly expanded or updated version of the cycle 2007 reading framework, with more research sources cited, including some published after the original framework. The corresponding section of the reading framework chapter in the cycle 2019 assessment framework document (CMEC, 2020) is mostly unchanged from the cycle 2016 document.

Neither the mathematics nor the science frameworks indicates that it is directly informed by education research in the respective content area. (They may be indirectly informed by research, however, through other frameworks consulted.)

Role of Other Frameworks

We identified a third kind of source informing assessment frameworks: Other frameworks for curricula or assessments.

The domain description sections of the different versions of the PCAP reading frameworks (those in the cycle 2007, cycle 2016, and cycle 2019 framework documents) do not reference any such sources.

By contrast, the mathematics framework indicates that it is based on (the assessment frameworks for) the School Achievement Indicators Program (SAIP, which preceded PCAP), PISA and TIMSS. The documents indicate it has been guided by two National Council of Teachers of Mathematics (NCTM) documents: *Principles and Standards for School Mathematics* and *Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics*. Although these different frameworks are described in the domain description section of the PCAP mathematics framework, their connection to the latter is not made explicit. That is, the PCAP mathematics framework does not report how its categories relate to the categories in these other frameworks.

The PCAP science framework also references the SAIP assessment framework and indicates it "takes into account findings from" PISA and TIMSS. (CMEC, n.d. c). However, the document seems to draw most heavily from another CMEC-authored framework, *Common Framework of Science Learning Outcomes K to 12* (CMEC, 1997).

Role of Professional Standards

The PCAP sources do not reference professional standards.

Sources for the Assessment Design

By "assessment design," we mean the way in which a domain description is made operational through weighting, test blueprints, item format decisions, and related specifications. The PCAP sources do not reference a process or other sources that inform the assessment design portion of the PCAP frameworks.

Authorship of Framework Documents

The first PCAP public report (from the 2007 cycle) indicates that in August 2003, a working group of "of experienced and knowledgeable representatives from several jurisdictions and including an external authority on measurement theory, large-scale assessment, and educational policy" (CMEC, 2008, p. 2) started the process of developing the assessment program. A "concept paper" (not cited) "would elaborate on issues of structure, development planning, operations, and reporting." (p. 2) The working group drew on this concept paper to "define" PCAP as follows:

"[PCAP will] be administered at regular intervals[,] be administered to students who are 13-year-olds at the start of the school year[,] be based on the commonality of all current juristictional [sic] curricular outcomes across Canada[,] assess reading, mathematics, and science[,] provide a major assessment of one domain with a minor concentration on the two other domains[, and] focus on reading as the major domain in the first administration in 2007. For each subject area, a thorough review of curricula, current assessment practices, and research literature was then undertaken and reports were written to indicate the common expectations among all jurisdictions." (p. 2)

The sources do not document the membership of this group, nor reference working groups or identify authors of the individual subject-area frameworks.

The cycle 2016 technical report references a working group and a specific contractor for updating the reading framework, but not the composition of the group.

Constituency Review Processes

Program documentation does not reference external or public review of frameworks.

Processes for Reviewing, Updating, and Revising Existing Frameworks

The cycle 2016 technical report indicates that the reading framework was updated for that assessment year. The text does not specify a process for arriving at a decision to review or update the framework. The description of the revision process is brief and does not document directives or parameters for the update nor consensus or constituency review processes. The document does not describe the specific changes made to the reading framework. (These changes, however, can be assessed through document comparison.)

Approval

PCAP program documentation does not reference a formal approval process for frameworks.

Discussion of NAEP Reading Assessment and Implications of Proposed Framework

The Governing Board has been discussing proposed updates to the NAEP Reading Framework since last summer. Some of the more recent discussions and questions have focused on the need to better understand the nature and rationale for features that are already included in the current reading assessment.

COSDAM Chair Gregory Cizek and Vice Chair Carey Wright determined that it would be useful for Committee members to have an opportunity to engage in discussion on this topic at the upcoming COSDAM meeting. To support this discussion, they generated a list of questions and requested that the National Center for Education Statistics (NCES) provide written responses in advance of the COSDAM meeting.

The attached materials include: 1) an Overview document produced by NCES as background for discussions during the May quarterly meeting on the proposed NAEP Reading Framework update, and 2) the specific questions and answers produced to inform this COSDAM discussion. The materials for the May quarterly meeting include additional documents to support the full Board discussions on the proposed NAEP Reading Framework update.

Overview of the NAEP Reading Assessment and Projections

April 23, 2021

This document has been prepared in response to questions from the National Assessment Governing Board membership and staff regarding the current operational NAEP Reading Assessment in relation to the most recent draft of the 2026 updated framework. This document has three parts:

I. **Description of the Current Operational NAEP Reading Assessment in Relation to the Most Recent Draft of the 2026 Updated Reading Framework**
II. **Support Features, Relevant Research, and Development Processes in the Current Operational NAEP Reading Assessment**
III. **Implementing the Updated Framework and Maintaining Trend**

I. DESCRIPTION OF THE CURRENT OPERATIONAL NAEP READING ASSESSMENT IN RELATION TO THE MOST RECENT DRAFT OF THE 2026 UPDATED READING FRAMEWORK

Starting with the 1992 NAEP Reading Framework, a driving principle for the NAEP Reading assessment has been authenticity as a means of establishing face validity. Authenticity in the context of the NAEP reading assessment means that, to the extent possible, the assessment should reflect the reading experiences of students outside of the testing context. For example, the 1992 NAEP reading assessment was one of the first large-scale assessments to use only full-length, naturally occurring texts as its stimulus reading materials. The move to digital assessment under the current framework has allowed the NAEP reading assessment to reflect the digital reading experiences students encounter on a daily basis both inside and outside of school contexts. The draft 2026 updated framework continues to reflect the principle of authenticity.

Definition

The current NAEP Reading Framework lists the following definition: “The NAEP Reading Assessment is guided by a definition of reading that reflects scientific research, draws on multiple sources, and conceptualizes reading as a dynamic cognitive process.” This definition

applies to the assessment of reading achievement on NAEP and states that reading is an active and complex process that involves:

- Understanding written text;
- Developing and interpreting meaning; and
- Using meaning as appropriate to type of text, purpose, and situation.

The draft 2026 updated framework maintains the current construct of reading comprehension while expanding the definition to include, "to explicitly acknowledge the sociocognitive processes of reading. Reading comprehension is defined as making meaning with text and four key features are highlighted—contexts, readers, texts, and activities." More specifically, the draft 2026 framework says the following:

"Reading comprehension is making meaning with text, a complex cognitive process shaped by students' social and cultural influences. To comprehend, readers:

- Engage with text in print and multimodal forms;
- Employ personal resources that include foundational reading skills, language, knowledge, and motivations;
- Extract, construct, integrate, critique, and apply meaning in activities across a range of contexts."

Testing Experience

The NAEP reading assessment transitioned from a paper-based assessment (PBA) to a digitally-based assessment (DBA) in 2017. In the most recent DBA in 2019, each student's assessment session began with a tutorial that included student interactions with the tools and interface and concluded with a 3-minute practice session. Following the tutorial and practice session, students worked through two 30-minute cognitive blocks. The second block was followed by a 15-minute survey questionnaire.

Texts

In accordance with the 2019 NAEP Reading Framework, which was first implemented in 2009, there are two broad categories of passages that make up the NAEP reading assessment: literary and informational. Literary texts include fiction, literary non-fiction, and poetry. Informational texts include exposition, argumentation or persuasive texts, and procedural texts.

The draft 2026 framework calls for three types of texts—literature, social studies, and science—and the texts in the 2019 operational pool fall easily into these three categories.

Items

After the passages are reviewed and approved by the Governing Board, items are written to assess three ***cognitive*** targets under the current framework. The current framework specifies the three cognitive targets as: Locate/Recall, Integrate/Interpret, and Critique/Evaluate.

The draft 2026 updated framework proposes four ***comprehension*** targets: Locate/Recall, Integrate/Interpret, Analyze/Evaluate, and Use and Apply. The addition of Use and Apply addresses the need to assess students' ability to apply the understandings they have gained from interacting with the stimulus materials for a given purpose (e.g., preparing a page of a website or writing a message to the school board).

The current NAEP Reading Framework calls for the following **item types**:

- **Selected response** – This item type encompasses traditional, single-answer, multiple-choice items as well as more complex items that require multiple selections to be answered correctly. NAEP's shift to digitally-based assessment allowed for the introduction of technology-enhanced items, which include matching (drag and drop), grid, and select-in-passage items. Most selected response items are scored dichotomously (correct or incorrect), but more complex selected response items may be scored for partial credit.
- **Constructed response, short and extended** – This item type requires students to generate a written response. Short constructed response items can be answered with a few words or sentences and extended constructed response items may elicit a short paragraph. These items are scored by humans, using a scoring rubric. Short constructed response items are scored with 2- or 3-point rubrics. Extended constructed response items use a 4-point rubric.

Percentages of each item type are specified in the framework for each grade. Typically, NAEP reading blocks include one extended constructed response item, three to five short constructed response items, and three to seven selected response items. The typical NAEP reading block includes a total of 9–11 items.

The draft 2026 updated framework recommends continuing with these item types and provides percentage ranges for selected response, short constructed response, and extended constructed response items. The draft framework also encourages the continued use and exploration of technology enhanced item types.

Reporting

Results of the NAEP reading assessment are reported on a 0–500-point scale. Three scores are reported at each grade level: a composite, or overall reading score, and two sub-scale scores, one for literary texts and one for informational texts. The draft 2026 updated framework maintains the 0–500-point scale and recommends reporting at each grade level: a composite score and three sub-scale scores—reading to engage in literature, reading to engage in science, and reading to engage in social studies contexts.

NAEP Contextual Questionnaire Items

Following the completion of two cognitive blocks, students respond to a 15-minute survey questionnaire. There are two sections to the Contextual Questionnaire: Core and Reading-specific. Core survey items collect information on students' demographic characteristics, opportunities to learn in and out of the classroom, and educational experiences.

Reading-specific survey items focus on reading-related activities and experiences in and out of school. These items are designed to inform interpretations of the results.

In addition to the student questionnaires, teachers and administrators in schools that participate in NAEP also complete their own NAEP Questionnaires.

The draft 2026 updated framework maintains the current approach to the survey questionnaires along with recommendations for changes to the specific items in the reading surveys.

Assembling the NAEP Assessment Via Assessment Blocks

Each NAEP reading assessment is administered in two 30-minute assessment blocks, followed by a 15-minute block of contextual items. Although each student sees only two blocks, there are multiple blocks in each operational assessment as shown in the chart below. Matrix sampling of students and blocks enables NAEP to cover a broad range of content, while also minimizing the burden for students and schools.

Table 1a summarizes the number of NAEP reading assessment blocks administered in the 2019 operational assessment for grades 4, 8, and 12. Typically, each block contains 9–11 items.

Table 1a. 2019 Operational NAEP Reading Assessment Blocks and Item Pool

	Grade 4	Grade 8	Grade 12
Total Number of Blocks	12	15	15
Total Number of Items	118	149	132

Types of Assessment Blocks

Currently, two types of blocks make up the NAEP operational reading assessment: discrete blocks and scenario-based task blocks.

Discrete item (DI) blocks provide general instructions for students to read the passage and provide answers to each assessment item relating to the passages that are presented. All texts and all items are always available for student access and use. The current operational pool of DI blocks is comprised of both transadapted and newly developed blocks as described below.

- **Transadapted blocks** – These blocks are digital renditions of the assessment blocks used in the paper and pencil era of NAEP. These DI blocks make up about two-thirds of the current operational assessment.
- **Newly developed blocks** – These blocks were developed specifically for a digital platform. To take full advantage of the digital format, some of these blocks use print texts and some use texts that are “digitally native” and multi-modal. Some passages

contain embedded hyperlinks and videos. (Note that videos are not used as introductions to texts.) Items addressing video content do so in relation to passage content.

Scenario-based Task (SBT) blocks use both print and digitally native, multi-modal texts. In contrast with DI blocks, students can only access texts and questions sequentially, as the SBTs control the order in which students read texts and items and respond to questions. In this way, students are presented with sources and stimulus materials as needed to respond to items. Videos appear both as part of the texts that students read and as additional content but are not used as introductions to texts. Items addressing video information always do so in relation to the written text.

Table 1b summarizes the number of NAEP reading assessment discrete and scenario-based blocks in the current operational assessment for grades 4, 8, and 12.

Table 1b. 2019 Operational NAEP Reading Assessment Discrete and Scenario-Based Blocks

Block Type	Grade 4	Grade 8	Grade 12
Scenario-based Task Blocks	2	2	2
Discrete Blocks (Transadapted)	7	10	11
Discrete Blocks (Newly developed for DBA)	3	3	2
Total	**12**	**15**	**15**

II. SUPPORT FEATURES, RELEVANT RESEARCH, AND DEVELOPMENT PROCESSES IN THE CURRENT NAEP OPERATIONAL READING ASSESSMENT

Consistent with the principle of authenticity, the current operational NAEP reading assessment uses ***support features***, referred to as Universal Design Elements (UDEs) in the draft framework, that are intended to replicate the types of supports provided during reading instruction and practice in school and at home. One central principle is worth emphasizing: ***all*** *support features used in a particular block are available to* ***all*** *students who take that block.*

The types of support features available on the 2019 NAEP reading operational assessment include:

- Look-back buttons
- Pop-up notes
- Passage introductions
- Eliminate answer choice
- Highlighting and notetaking
- Text-to-speech on directions
- Zoom & selection of color themes
- Multi-part response frames

- Purpose statements*
- Avatars
- Graphic organizers
- Item foreshadowing
- Directions and transitions
- Item resetting

* Purpose statements are not considered UDEs in the draft 2026 updated framework.

Not all features are available in every block, but all of the current operational NAEP reading blocks include some support features. Some of these features are available for all reading blocks, and across other NAEP subjects, at the system level (e.g., highlighting, text-to-speech on directions, zoom, and color themes). Others are content-specific, including item look-back buttons, pop-up notes, passage introductions, and multi-part response frames (complex items with multiple components split into multiple response fields). Others appear only in SBTs, or a subset of SBTs, depending on the goals of the tasks, including block-specific purpose statements, avatars, graphic organizers and sequential directions and transitions.

The following subsection provides additional information about the use of pre-reading features, pop-up notes, and avatars and pop-up notes.

Pre-Reading Features

The current operational assessment includes two types of pre-reading features: ***block-specific purpose setting statements*** and ***introductions*** to specific texts, which have been developed in consultation with the Reading Standing Committee[1] and approved by the Assessment Development Committee (ADC) on behalf of the Governing Board. The current NAEP Reading Framework does not provide guidance on pre-reading features.

Purpose Setting

DI blocks include general directions to "read and answer the questions," but do not include block-specific purpose statements.

SBTs include both general directions and block-specific purpose statements. Block-specific purpose statements introduce a purpose for reading and describe the task students are to complete (e.g., gather information for a webpage or to compose an email message). Block-specific purpose statements focus on the tasks students will perform rather than on introducing specific texts. Block-specific purpose setting statements appear in six of the blocks (2 per grade) in the 2019 operational reading pool (17% of the pool).

Introductions

A small number of DI blocks include some information about the text students are about to read prior to reading. This prereading feature has appeared in NAEP Reading since before the

[1] The Reading Standing Committee is a diverse group of experts and state assessment staff in reading from across the nation. They advise as part of the assessment item development process, ensuring that NAEP assessment items align to the NAEP framework. There is a Standing Committee for each NAEP subject area assessment.

digitally-based assessment began in 2017. All introductions are written text; none are multi-media (video or audio) as was proposed in the draft 2026 framework.[2]

Passage-specific introductions appear in eight of the blocks across all three grades in the 2019 operational reading pool (23% of the pool). Five of these introductions were added by the test developers and three were part of the original source. In five of these instances, the introduction provides some information about the author. In two of these instances, the introduction provides context for passages that are excerpts.

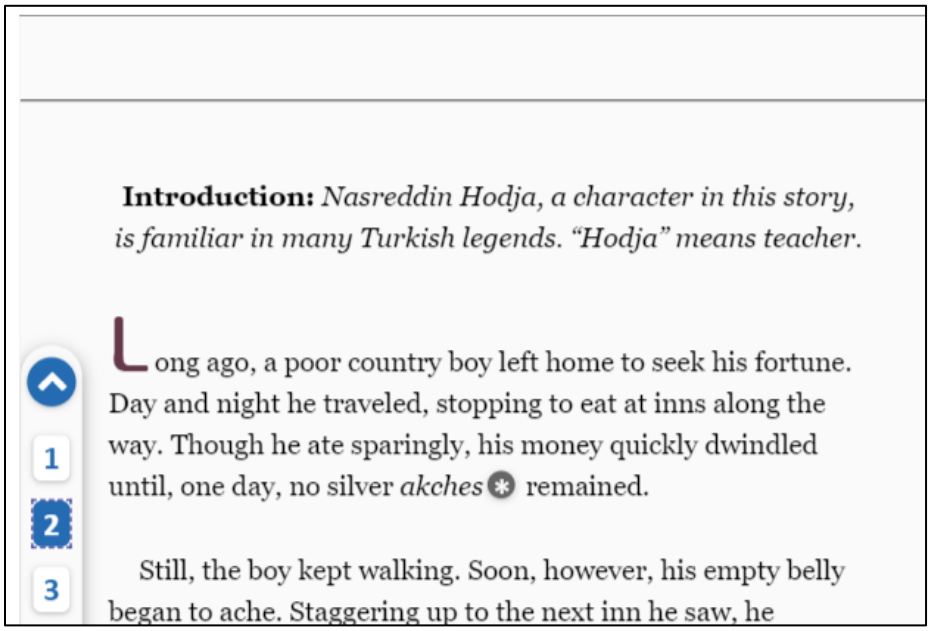
Generally, there are no consensus assessment industry guidelines or standards for when/how to provide introductions, though there is an extensive research base on the role of prior knowledge in reading comprehension that provides some guidance. For example, seminal research on schema theory by John Bransford and his colleagues found that readers were only able to adequately demonstrate their reading comprehension skills with passages written in general terms when titles were provided that activated their schema/prior knowledge about the topics of the passages. This work, along with content analyses of instructional materials and cognitive labs with students, enabled NCES to implement passage introductions in the operational NAEP reading assessment.

In addition, introductions were deemed important by the Reading Standing Committee as a means of orienting the reader and as a response to the need for content and face validity evidence. In timed, on-demand assessments such as NAEP, brief framing can help to mitigate construct-irrelevant variance, and such introductions and framings are common in sources students encounter in their daily lives. Periodically, the NAEP program invites all states and participating Trial Urban District Assessment (TUDA) districts to review the entire pool of NAEP items. The most recent state/TUDA item review in 2015 included texts with introductions, and no concerns were raised regarding these features. Finally, text introductions appear in some state reading assessments, such as PARCC and Smarter Balanced.

The following examples of passage introductions from previous NAEP reading assessments come from released and publicly available blocks (i.e., not the current operational pool). The first is an introduction to a Turkish folktale called “Five Boiled Eggs.” The second introduces an article about the writer, E. B. White, and the third introduces an essay by E. B. White, by explaining that the author of the essay they are about to read is also a children’s author. The E.B. White passages appeared in the paper assessment and were released in 2011. The “Five Boiled Eggs” passage appeared in the paper assessment and was transadapted for the digital assessment in 2017 and released after that administration.

[2] Responding to the Governing Board’s March 2021 Board meeting deliberations, the April 2021 draft of the 2026 framework update does not include multimedia introductions.

Example 1.



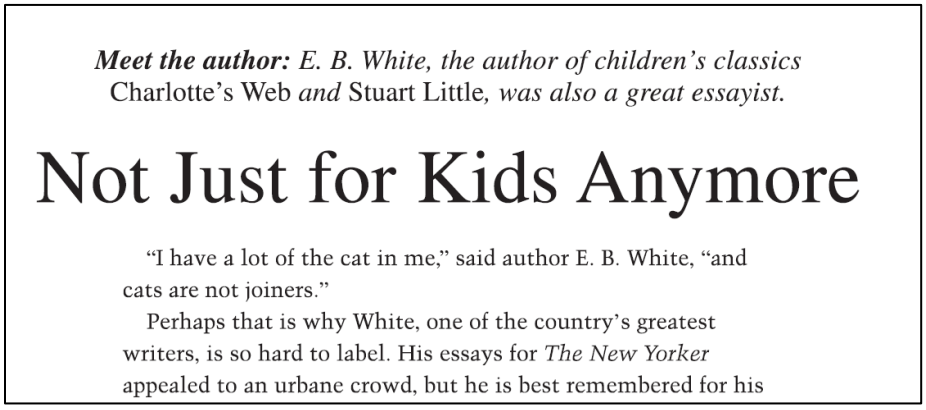
Introduction: *Nasreddin Hodja, a character in this story, is familiar in many Turkish legends. "Hodja" means teacher.*

1
2
3

Long ago, a poor country boy left home to seek his fortune. Day and night he traveled, stopping to eat at inns along the way. Though he ate sparingly, his money quickly dwindled until, one day, no silver *akches* remained.

Still, the boy kept walking. Soon, however, his empty belly began to ache. Staggering up to the next inn he saw, he

Example 2.



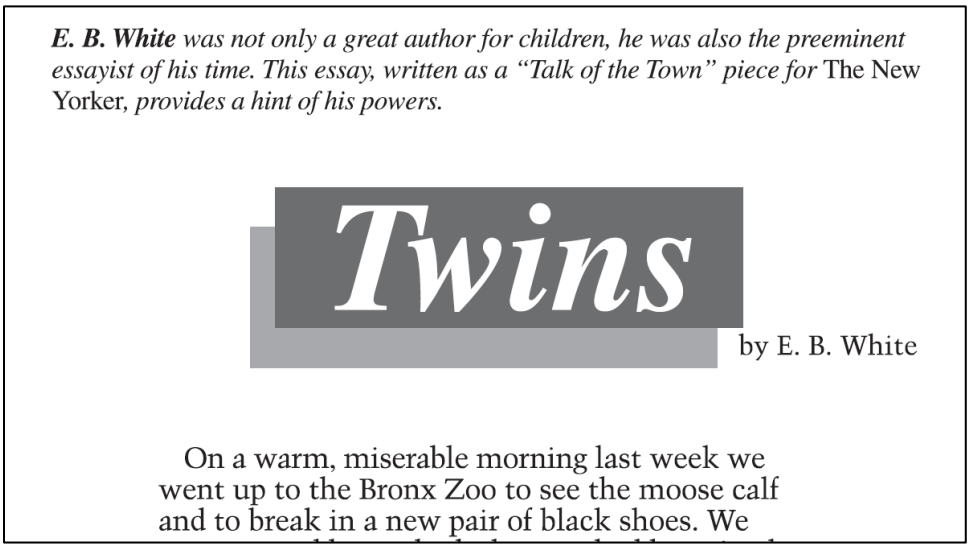
Meet the author: *E. B. White, the author of children's classics* Charlotte's Web *and* Stuart Little*, was also a great essayist.*

Not Just for Kids Anymore

"I have a lot of the cat in me," said author E. B. White, "and cats are not joiners."

Perhaps that is why White, one of the country's greatest writers, is so hard to label. His essays for *The New Yorker* appealed to an urbane crowd, but he is best remembered for his

Example 3.



E. B. White *was not only a great author for children, he was also the preeminent essayist of his time. This essay, written as a "Talk of the Town" piece for* The New Yorker, *provides a hint of his powers.*

Twins

by E. B. White

On a warm, miserable morning last week we went up to the Bronx Zoo to see the moose calf and to break in a new pair of black shoes. We

Avatars

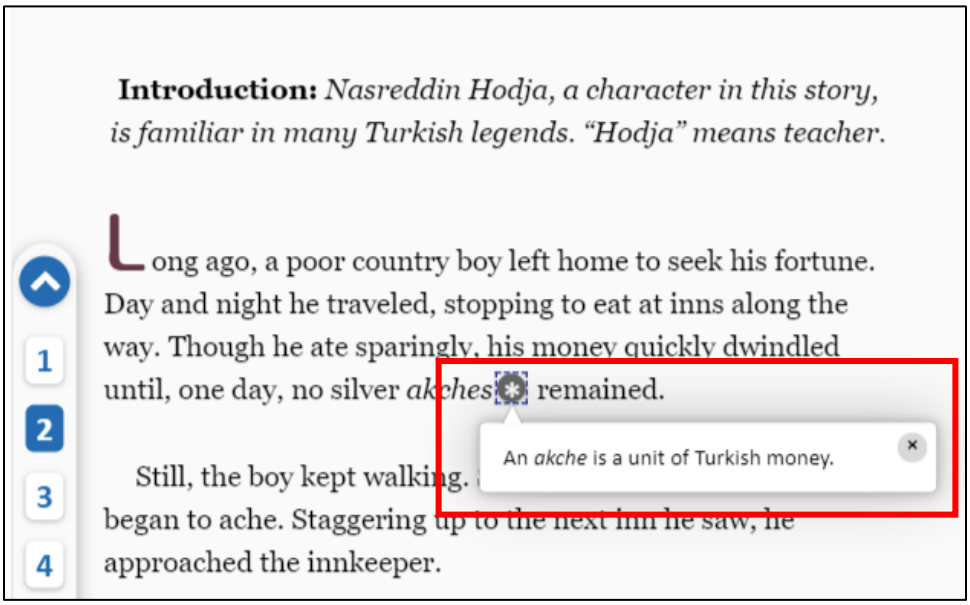
Avatars are task characters used to create a social context and facilitate purpose-setting and transitions in SBT blocks (no discrete blocks use them) but are in and of themselves neither purpose statements nor introductions. Two of the total pool of six SBT blocks, across grades, in the 2019 operational reading pool use avatars (6% of the blocks in the total pool).

Pop-up Notes

Pop-ups are indicated by buttons in the text that signal to students that they can read more about a word or phrase. These kinds of notes appeared on the paper-based assessment (PBA) as footnotes. Pop-up notes occur in three blocks in the 2019 operational pool (9% of the blocks in the total pool). Two of the three pop-up instances provide definitions of words/terms that may be unfamiliar to the reader and are important to overall understanding. The third instance presents information that was provided in the original text. There are no assessment items directly related to the information in the pop-up notes.

The following example shows a pop-up note from the passage "Five Boiled Eggs."

Example 1.



A substantial proportion, 63%, of the entire pool of reading blocks in the 2019 operational assessment *does not contain the pre-reading features* described above. These blocks could be characterized as providing opportunities for "cold reads" and will continue to be part of the operational assessment in 2026.

Relevant NAEP Research

As noted above, two types of blocks make up the NAEP operational reading assessment: Discrete Item (DI) blocks and Scenario-based Task (SBT) blocks. At each grade level in 2019 (as noted in Table 1b above), two of 12 grade 4 blocks are SBTs, two of 15 grade 8 blocks are SBTs, and two of 15 grade 12 blocks are SBTs. The remainder are DI blocks. A special study was conducted in 2018 to examine the SBT format, relative to the current framework. For this study, researchers created discrete versions of reading SBT blocks using the same texts and items for both versions. This special study compared student performance on the same set of items and passages in a DI block versus an SBT block.

Although this study was conducted before the framework update project began, it is relevant to conversations about the framework update because SBTs involve collections of support features, which are referred to in the framework update as Universal Design Elements (UDEs). Both SBT blocks and DI blocks include UDEs.

Three of the 15 UDEs in the draft 2026 updated framework only appear in SBT blocks (i.e., avatars, sequential directions and transitions, and item resetting). The remaining 12 UDEs,

including text introductions and pop-up notes, can appear in either SBT or DI blocks.[3] This study provides no information about specific UDEs. Instead, the study examines collections of UDEs in an SBT format.

This was a randomized control trial study with a total of 3,000 students, counterbalanced for version, genre (literary and informational), and block position at each grade. Both the SBT and DI versions of blocks were delivered on tablets. Consistent with the students' experience with DI and SBT blocks in the operational assessment, students were able to move among texts and items at will in the discrete version, whereas movement between texts and items was sequential in the SBT versions.

Key findings (The differences summarized below are statistically significant.):

- Students taking the SBT versions of blocks outperformed students taking the DI versions of block in four of the six blocks.
- The advantage for the SBT versions was consistent across all NAEP subgroups (gender, race, SES, disability, ELL). In other words, there is no differential effect for any subgroup.
- The advantage of the versions with support features was consistent for low- and high-performing students in four of the six blocks.
- For the four blocks for which performance on the SBT version was significantly higher, the differences in percent correct ranged between 2% and 8%, with an average of 5%.
- The SBT-DI special study provides some indication that SBT versions of items tend to be more engaging/motivating to students than DI versions of items. This tendency could contribute to students' higher performance on SBT versions of items, compared with DI versions of items.
- Generally, reading SBT blocks tend to be equally or more difficult than DI blocks, but when comparing SBT and DI versions of the same set of items, SBTs tend to be less difficult than their DI versions.[4]
- Speededness was more of an issue in SBT versions. Revisions were made to reduce speededness before these blocks became part of the operational assessment.

[3] Of the 15 UDEs listed in the February 26, 2021 draft of the reading framework update, 13 already appear in the reading assessment. The 2 additions would have been: student exemplars as mentor texts (a task-based UDE) and multimedia passage introductions (a knowledge-based UDE). However, multimedia passage introductions were removed from the latest draft of the framework update. Text introductions already appear on the assessment – see earlier sections of this document on (1) pre-reading features and (2) existence of "cold reads".

[4] Because NAEP uses an Item Response Theory (IRT) model to generate scores, adding more difficult items to the NAEP Reading Item Pool will improve measurement at the high end of the score scale, i.e., detect smaller differences in student achievement for higher performers. Conversely, adding less difficult items will improve measurement on the low end of the score scale, i.e., detect smaller differences in student achievement for lower performers. The IRT methodology for scoring ensures that adding harder items to the item pool will not artificially lower scores and that adding easier items to the item pool will not artificially inflate scores.

Other Standard Research and Reviews in NAEP Item Development

NCES implements a routine research and development cycle to develop every assessment block carefully before it is introduced to the operational NAEP assessment. Each new block undergoes systematic scrutiny, typically including these steps:

1. **Text Selection.** Texts and text sets are identified by the ETS reading item development team at a rate of four for every one text or text set expected to become part of the operational assessment. Proposed texts are reviewed by the ETS bias and sensitivity review team and the ETS editorial staff and are ultimately reviewed and approved by the NCES item development staff and contractors, and the Governing Board Assessment Development Committee (ADC).
2. **Initial Item Reviews.** After passages are approved, items are developed by the ETS reading item development team. Once draft items are completed, ETS reviewers conduct editorial, cold read, bias and sensitivity, and language accessibility reviews. They are then reviewed by NCES item development staff and contractors and the reading standing committee.
3. **Pretesting.** Following initial item review, items and support features are pretested, using:
    a. Cognitive interviews with individual students to determine how they respond to proposed new texts and comprehension test items. The purpose is to determine whether the tasks actually engage students in the intended comprehension processes.
    b. Tryouts under "live" testing conditions with 50–200 students from the target population to determine whether a wide range of students can complete the blocks within the allocated time and whether all of the parts of the block are working as intended.
    c. Usability studies, which test new item or passage interactions with small groups of students.
4. **Revised Item Reviews**. After items are pretested and revised by item developers, ETS reviewers conduct editorial, bias and sensitivity, and language accessibility reviews. They are then reviewed by NCES item development staff and contractors and the standing committee. Item revisions are adjudicated with NCES item development staff, and items are submitted to the Governing Board ADC for final review and clearance for piloting. Before piloting, state/TUDA reviews may occur.
5. **Piloting.** Proposed new blocks are folded into the administration of operational blocks of a live assessment. By comparing student and item performance across the new and the old blocks, NAEP developers can determine whether the new blocks effectively scale together with the old, measuring the same underlying comprehension construct.

6. **Post-pilot Reviews**. Following the collection of pilot data (n=2500–3000 students per form), the following groups review pilot data, item level analyses, texts, and items:
    - ETS reading item development team
    - ETS data analysis and reporting team
    - ETS Differential Item Functioning (DIF) panel
    - ETS bias and sensitivity review team
    - NCES item development staff and contractors
    - NCES data analysis and reporting staff and contractors
    - ETS editorial staff
    - ETS Reading Standing Committee
    - Governing Board Assessment Development Committee (ADC)

III. IMPLEMENTING THE UPDATED FRAMEWORK AND MAINTAINING TREND

This section provides information about the implementation of the updated framework and is based on the contents of the latest draft of that document.

Following Board adoption of an updated framework, it will take time to develop the assessment. As new content is piloted and approved, old content, in particular blocks transadapted from the paper-based assessment, can be phased out. Most importantly, this gradual item development for the updated framework allows for trend to be maintained.

The 2022 and 2024 assessments will be the last operational assessments that are fully aligned to the current framework. The 2026 assessment is projected to be the first operational assessment under the updated framework. The 2026 assessment would include both trend blocks from the 2022 and 2024 operational assessments and newly developed blocks piloted in 2024, being used for the first time in an operational assessment.

In the Governing Board's previous discussions of the updated framework, concerns were expressed that there would be insufficient carryover of content to maintain trend in 2026. However, the information below indicates that maintaining trend in 2026 is possible with careful planning. The projected contents of the next three operational assessments are as follows:

- 2022:
    - Grades 4 and 8 – trend content only (re-administration of 2019)
    - Grade 12 – no assessment
- 2024:
    - Grades 4 and 8 – trend content (all blocks carried over from 2022) plus new operational content (drawn from blocks piloted in 2017 and 2019)
    - Grade 12 – trend content only (re-administration of 2019)

- 2026:
    - Grades 4 and 8 – trend content (all blocks carried over from 2024) and new operational content (drawn from blocks piloted in 2024)
    - Grade 12 – no assessment

Projected Numbers of Blocks Available for the 2026 Operational Reading Assessment

The tables below include information about the numbers of blocks in each of the following two categories that will make up the 2026 operational assessment.

1. **Trend blocks**, which consist of discrete blocks from the 2022 and 2024 operational assessments, which do not include block-specific purpose statements, and SBT blocks from the 2024 operational assessments, which do include block-specific purpose statements.
2. **New operational blocks** developed to address new aspects of the updated framework, including block-specific purpose statements and the updated comprehension targets. These blocks are being used for the first time in the 2026 operational assessment and will not become trend blocks until they are administered operationally for the second time.

The proposed approach to a gradual implementation of the updated framework has been revised since the original Overview document was submitted to Governing Board staff just prior to the March 2021 Board meeting. The March 2021 version of this document suggested adding block-specific purpose statements to three existing discrete blocks at each grade and re-piloting them in 2024. However, the most recent plan retains the existing discrete blocks, as is, and redirects the funds that would have been used for modification and re-piloting of existing blocks to the development of new blocks under the aegis of the updated framework. The current plan provides for a carryover of blocks from the 2024 to the 2026 assessment of 80% at grade 4 and 83% at grade 8. **Although ultimately an empirical question, these percentages of carryover should allow for the maintenance of trend.**[5] (See Table 3a below.)

The current plan for the 2026 development proposes new pilot development of six blocks at grades 4 and 8[6] to yield four new operational blocks. It also assumes that blocks piloted in 2017, 2019, and 2024 will be approved for operational use and that there are no public releases prior to the 2026 assessment.

Tables 3a, 3b, and 3c provide information about the composition of the 2026 operational assessment based on the current plan.

[5] The current NAEP reading framework – adopted in 2004 and first implemented in 2009 – included no carryover from the previous framework (0 percent) and trend was maintained. To learn more about how trend was maintained for the 2009 NAEP Reading Assessment, see the Reading Trend Study description at https://nces.ed.gov/nationsreportcard/reading/trend_study.asp.

[6] Grade 12 will not be administered in 2026 and new grade 12 development is out of scope.

Table 3a. Projected Numbers of Blocks by Status available for the 2026 NAEP Operational Reading Assessment at Grades 4 and 8

Blocks	Grade 4	Grade 8
TREND	16 (80%)	20 (83%)
NEW OPERATIONAL	4	4
Total Blocks	**20**	**24**

As a result of needing to both maintain trend and introduce new content aligned with the updated framework, the 2026 operational assessment is projected to include more blocks at each grade than the 2022 operational assessment. The grade 4 assessment would contain 11 blocks in 2022[7] and as many as 20 in 2026, and the grade 8 assessment would contain 14 blocks in 2022 and as many as 24 blocks in 2026.[8] A larger item pool is also required to support reporting goals for the updated framework, including reporting for three subscales instead of the two subscales reported under the current framework.

All of the passages and items in the blocks that would be carried over from 2024 to 2026 are consistent with the updated framework. The block-specific purposes required by the updated framework will be present in 40% of the blocks at grade 4 and 33% of the blocks at grade 8.

Tables 3b and 3c describe the contents of the projected 2026 operational assessment at each grade broken down by subscale.

[7] Tables 1a and 1b showed that the 2019 assessment included 12 blocks at grade 4 and 15 blocks at grade 8. However, one cross-grade 4/8 block has been dropped for sensitivity reasons, resulting in 11 blocks at grade 4 and 14 blocks at grade 8 for the 2022 assessment. Blocks sometimes need to be dropped for sensitivity reasons if they address topics that might be disturbing because of recent or ongoing current events, e.g., a hurricane, a pandemic, etc.

[8] The actual number of 2026 blocks is contingent on the contents of possible public releases in 2022 and 2024.

Table 3b. Projected Number of Blocks available for the 2026 NAEP Operational Reading Assessment by Status and Subscale at Grade 4

Blocks	Reading in Literature	Reading in Social Studies	Reading in Science	Total Blocks
TREND	7	4	5	16 (80%)
NEW OPERATIONAL	*New development would include at least one block in each of the reading in social studies and science contexts.*			4
Total Blocks				**20**

Table 3c. Projected Number of Blocks available for the 2026 NAEP Operational Reading Assessment by Status and Subscale at Grade 8

Blocks	Reading in Literature	Reading in Social Studies	Reading in Science	Total Blocks
TREND	8	6	6	20 (83%)
NEW OPERATIONAL	*New development would include at least one reading in literature block.*			4
Total Blocks				**24**

Appendices 1 and 2 on the following pages depict the movement of blocks across the 2022, 2024, and 2026 assessments at grades 4 and 8, as well as the addition of newly developed blocks.

Appendix 1. Proposed Composition of the 2022, 2024, and 2026 Assessments at Grade 4 by Context and Status

2022 Assessment	2024 Assessment	2026 Assessment
Rdg in Lit Block 1	Rdg in Lit Block 1	Rdg in Lit Block 1
Rdg in Lit Block 2	Rdg in Lit Block 2	Rdg in Lit Block 2
Rdg in Lit Block 3	Rdg in Lit Block 3	Rdg in Lit Block 3
Rdg in Lit Block 4	Rdg in Lit Block 4	Rdg in Lit Block 4
Rdg in Lit Block 5	Rdg in Lit Block 5	Rdg in Lit Block 5
	Rdg in Lit Block 6	Rdg in Lit Block 6
	Rdg in Lit Block 7	Rdg in Lit Block 7
Rdg in Science Block 1	Rdg in Science Block 1	Rdg in Science Block 1
Rdg in Science Block 2	Rdg in Science Block 2	Rdg in Science Block 2
Rdg in Science Block 3	Rdg in Science Block 3	Rdg in Science Block 3
Rdg in Science Block 4	Rdg in Science Block 4	Rdg in Science Block 4
	Rdg in Science Block 5	Rdg in Science Block 5
Rdg in SocSt Block 1	Rdg in SocSt Block 1	Rdg in SocSt Block 1
Rdg in SocSt Block 2	Rdg in SocSt Block 2	Rdg in SocSt Block 2
	Rdg in SocSt Block 3	Rdg in SocSt Block 3
	Rdg in SocSt Block 4	Rdg in SocSt Block 4
	Pilot Block A	New Op Block
	Pilot Block B	New Op Block
	Pilot Block C	New Op Block
	Pilot Block D	New Op Block
	Pilot Block E	
	Pilot Block F	

KEY
Trend Block
New Operational Block
Pilot Block

Appendix 2. Proposed Composition of the 2022, 2024, and 2026 Assessments at Grade 8 by Context and Status

2022 Assessment	2024 Assessment	2026 Assessment
Rdg in Lit Block 1	Rdg in Lit Block 1	Rdg in Lit Block 1
Rdg in Lit Block 2	Rdg in Lit Block 2	Rdg in Lit Block 2
Rdg in Lit Block 3	Rdg in Lit Block 3	Rdg in Lit Block 3
Rdg in Lit Block 4	Rdg in Lit Block 4	Rdg in Lit Block 4
Rdg in Lit Block 5	Rdg in Lit Block 5	Rdg in Lit Block 5
Rdg in Lit Block 6	Rdg in Lit Block 6	Rdg in Lit Block 6
	Rdg in Lit Block 7	Rdg in Lit Block 7
	Rdg in Lit Block 8	Rdg in Lit Block 8
Rdg in Science Block 1	Rdg in Science Block 1	Rdg in Science Block 1
Rdg in Science Block 2	Rdg in Science Block 2	Rdg in Science Block 2
Rdg in Science Block 3	Rdg in Science Block 3	Rdg in Science Block 3
Rdg in Science Block 4	Rdg in Science Block 4	Rdg in Science Block 4
Rdg in Science Block 5	Rdg in Science Block 5	Rdg in Science Block 5
Rdg in SocSt Block 1	Rdg in Science Block 6	Rdg in Science Block 6
Rdg in SocSt Block 2	Rdg in SocSt Block 1	Rdg in SocSt Block 1
Rdg in SocSt Block 3	Rdg in SocSt Block 2	Rdg in SocSt Block 2
	Rdg in SocSt Block 3	Rdg in SocSt Block 3
	Rdg in SocSt Block 4	Rdg in SocSt Block 4
	Rdg in SocSt Block 5	Rdg in SocSt Block 5
	Rdg in SocSt Block 6	Rdg in SocSt Block 6
	Pilot Block A	New Op Block
	Pilot Block B	New Op Block
	Pilot Block C	New Op Block
	Pilot Block D	New Op Block
	Pilot Block E	
	Pilot Block F	

KEY
Trend Block
New Operational Block
Pilot Block

NCES Response to the Committee on Standards, Design and Methodology (COSDAM) Reading Framework Questions

April 22, 2021

This document has been prepared in response to questions from the National Assessment Governing Board's COSDAM regarding the current operational NAEP Reading Assessment in relation to the most recent draft of the 2026 updated framework. Three groups of questions are addressed in this document: 1) questions regarding Universal Design Elements (UDEs); 2) questions about the construct(s) being measured and the feasibility of maintaining trend; and 3) questions about implementation plans, projections, and budget considerations.

Universal Design Elements (UDEs) Questions

What research evidence was used to implement the "support features" on the assessment, in particular the passage introductions? Is there existing evidence that knowledge-based UDEs are differentially effective based on students' prior knowledge? Is there any existing evidence regarding the "effect size" of UDEs on performance?

UDEs, such as introductions, have been part of the NAEP Reading Assessment since before NAEP became a digital assessment in 2017. In general, there are no assessment industry guidelines or standards for when/how to provide introductions, though there is an extensive research base on the role of prior topic knowledge in reading comprehension that provides some guidance. For example, seminal research on schema theory by John Bransford and his colleagues found that readers were only able to adequately demonstrate their reading comprehension skills with passages written in general terms when titles were provided that served to activate their schema/prior knowledge about the topics of the passages. This work, along with content analyses of instructional materials and cognitive interviews with students, provided justification for NCES to implement passage introductions in the operational NAEP Reading Assessment. It is also the case that text introductions appear in some state reading assessments, such as the Partnership for Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced Assessment Consortium.

Examining the differential effectiveness of introductions and pop-up notes based on students' prior topic knowledge would require a study that includes measures of students' background knowledge. The NAEP program has not conducted any such study. Similarly, the NAEP program does not have evidence regarding the "effect size" of UDEs on performance because NAEP is not primarily a research program. NAEP relies on data from a variety of quantitative and qualitative sources to inform its development including cognitive interviews, small-scale tryouts, content reviews, and, occasionally, special studies. All new NAEP reading blocks are evaluated in a nationally representative pilot, followed by rigorous, block- and item-level analyses, and submitted for further review by the Governing Board's Assessment Development Committee (ADC). Weak or problematic blocks are not moved on for inclusion in the operational assessment.

Pretesting through cognitive interviews and small-scale tryouts was used to explore new UDEs introduced with Scenario-based Tasks (SBTs). The evidence from pretesting indicated that the majority of students reported that SBT UDEs were helpful and not distracting.

How much time do existing UDEs add to testing? Would additional UDEs exacerbate this further?

Is there any evidence about whether this additional time might hinder performance or be distracting?

The incorporation of UDEs in reading blocks is an integral part of the development of 30-minute blocks, as opposed to an "add on." Any potential time or cognitive burden they may pose is evaluated for each block as part of the development and pretesting processes via cognitive interviews and small-scale tryouts. Independent of the impact of UDEs, speededness is evaluated and addressed for all blocks as part of the development, pretesting, and piloting processes. The majority of the UDEs recommended in the draft framework are already included in the reading blocks in the NAEP operational assessment.

Is NCES concerned about the framework's characterization of "support features" as UDEs given how the NAEP program already characterizes "Universal Design Elements?"

NCES describes "Universal Design Elements" as a form of accommodation that is available to all students. Some of the UDEs described by NCES, such as highlighting and zoom, are considered "task-based UDEs" in the updated framework. NCES is not concerned that the updated framework adds UDEs that are not specified in its current description of UDEs.

What should be the main takeaways from the Scenario-based Task—Discrete Block (SBT-DI) study, relating to this framework update?

The main takeaway from the SBT-DI study is that regardless of ability (low vs. high), students performed better on the SBTs[1]. The main difference between the SBT and DI versions was the use of purpose-driven introductions and a broad range of UDEs in the SBTs. The support features in SBTs examined in the study are consistent with the updated framework.

Construct/Trend Questions

In NCES's view, does the current framework and the framework update both allow for "cold reads?"

The assessment has not been operationalized in terms of "cold reads" as the term is not defined or discussed in either the current or draft updated framework.

What evidence/ongoing studies/best guesses are there related to the likelihood of maintaining trend? Is this only a function of how many new blocks are needed? In NCES's view, do the newly proposed UDEs (mentor texts and multi-media introductions) represent a change in the construct of reading that can threaten trend (on a conceptual level)? In NCES's view, do other strictly digital UDEs from the framework represent a change in the construct of reading that should have already threatened trend (on a conceptual level)?

The likelihood of maintaining trend is a function of both how many new blocks are needed and whether these blocks differ qualitatively from the existing blocks in terms of what they measure. That said, there is a high likelihood of maintaining trend under the updated framework. Evidence of this comes from several sources. First, the construct of reading comprehension in the updated framework has changed very little from the construct in the current framework. This means that the passages and items developed under the updated framework will not differ significantly from those that were developed under the current framework. Second, the current plan of gradual implementation results in carryover of approximately 80% of blocks from the 2024 to the 2026 operational assessments. Finally, it should be noted that trend was maintained in the implementation of the current framework when there was no carryover from the previous assessment.

The possibility that the new UDEs recommended in the draft framework (mentor texts and multimodal introductions) could threaten the construct is also an empirical question that will be investigated through means such as pretesting and, possibly, special studies and evaluated

[1] The magnitude of the improvement, in terms of percent correct, ranged between 2% and 8% with an average of 5%. Although the SBT versions were less difficult than the DI versions of the same texts and items, evidence from the operational assessment indicates that SBTs are of equal or greater difficulty than the DI blocks in the operational pool.

through the various reviews described in the most recent memorandum—*Overview of the NAEP Reading Assessment and Projections*—to the ADC.

The "strictly digital" UDEs introduced with SBTs in 2019 are not a threat to trend either conceptually or empirically. Conceptually, they are consistent with the construct of reading comprehension in the current framework as a means of measuring students' ability to "use meaning as appropriate to type of text, purpose, and situation" (part of the definition from the current NAEP Reading Framework [2009]). Empirically, SBT blocks containing these UDEs scale with Discrete Blocks that do not include these UDEs.

Implementation and Budget Questions

To what extent can the current reading item pool be used to implement the framework update? Specifically, how much re-field testing is needed and how much new item development is needed? What is the cost of implementing the framework update?

NCES's proposed approach to a gradual implementation of the updated framework has been revised since the March 2021 Board meeting (see also in this packet of materials—*Overview of the NAEP Reading Assessment and Projections*). The current plan increases the percentage of trend blocks carried over from the 2024 to the 2026 assessments, to 80% at grade 4 and 83% at grade 8. Although ultimately an empirical question, these percentages of carryover should allow for the maintenance of trend.

All of the passages and items in the blocks that would be carried over from 2024 to 2026 are consistent with the updated framework. The block-specific purposes required by the updated framework will be present in 40% of the blocks at grade 4 and 33% of the blocks at grade 8. The remainder of trend blocks include general purpose statements.

NCES expects to develop and pilot six new blocks at each of grades 4 and 8 to yield four new operational blocks. It also assumes that blocks piloted in 2017, 2019, and 2024 are approved for operational use, and there are no public releases prior to the 2026 assessment.

The cost of implementing the updated framework can only be determined when an updated framework has been approved.

Age 17 NAEP Long-Term Trend (LTT) Assessments in 2022

Due to the school closures in response to COVID-19, the age 17 NAEP LTT mathematics and reading assessments, which were originally scheduled to be administered in 2020 along with the corresponding LTT assessments for ages 9 and 13, have been postponed to 2022 and will be administered between end of March and May 2022. All items to be administered in 2022 come from the 2012 operational assessments.

The reporting scales for both Mathematics and Reading LTT are univariate with no subscales reported. In the base year (i.e., 1973 for Mathematics and 1971 for Reading), a cross-age scale was established, which placed the results of all three age groups on the same scale. After the base year, however, the IRT scaling of three age groups has been conducted separately. In other words, the data analyses from one age group have no impact on those of the other age groups. Therefore, the two-year gap between the age 17 assessment and the assessment of the other two age groups does not pose an issue in terms of scaling. On the other hand, there are a few potential issues that are worth considering.

One potential psychometric concern of NCES is that the age 17 LTT assessments, if administered in 2022, will have a ten-year gap when trending back to 2012. That is two years longer than age 9 and age 13 LTTs. Because of this larger gap and possible learning loss due to the pandemic, NCES anticipates more age 17 trend items to function differently between 2022 and 2012 and hence being split[1] in scaling, as compared to LTTs of ages 9 and 13. As a reference, the table below summarizes the number of items treated[2] for Mathematics LTT and Reading LTT in 2012 and 2020. The 2012 LTTs were 4 years apart from the previous LTT assessments (2008) and very few trend items were split in scaling for both subjects of all three age groups. In contrast, both age 9 and age 13 Mathematics LTTs in 2020 had about 14 percent of the item pool (19 and 22 items respectively) split. At what point the number of items treated items become problematic is an empirical question that likely depends on the specific assessment and the nature of the remaining trend items in terms of how representative they are of the construct being measured.

[1] When a trend item is "split" between the two consecutive years being linked, it is treated as two different items, and hence, it no longer serves as a trend item.

[2] There are three kinds of treatments: an item might be split, or dropped from the analysis all together, or one or more score categories might be collapsed.

Table 1. 2012 & 2020 Item Treatment Summary for the NAEP Long-Term Trend Assessments

Subject	Age	Year	Number of Items	Number of Items Treated	Number of Items with Categories Collapsed	Number of Items Split	Number of Items Dropped
Mathematics	Age 9	2012	136	3	1	0	2
		2020	135	22	1	19	2
	Age 13	2012	157	6	0	2	4
		2020	152	25	0	22	3
	Age 17	2012	155	5	0	2	3
Reading	Age 9	2012	88	3	3	1	0
		2020	78	2	2	0	0
	Age 13	2012	106	4	1	3	1
		2020	95	5	1	5	0
	Age 17	2012	103	2	0	2	0

Note: Details may not add up to total because some items had more than one type of treatment.

Another important factor is the degree to which the pandemic continues in 2022. If students take the LTT assessments under COVID-19 mitigation protocols in 2022, changes to the normal, pre-pandemic administration procedures may adversely affect the students' performance. This would introduce a potential confounding variable that will make the interpretation of the trend results for this age group challenging. Needless to say, this issue is not unique to age 17 LTT, but relevant to all assessments in 2022. In addition, given the likely learning losses due the ongoing pandemic, the results of this assessment might show scores for age 17 in 2022 that in the score range of age 13 cohort in 2020. This would render the interpretation of the results for this group very challenging. Note that, regardless of the actual results, the report card for the 2022 age 17 would not include the performance of the other two ages assessed in 2020.

A final issue is related to the utility of the age 17 LTT PBA data point in 2022. LTT assessments, originally scheduled for 2020 for all age groups, were supposed to provide an additional data point under the paper-based assessment (PBA) condition before the assessments are transitioned to DBA in 2024. Given the potential psychometric and validity issues described above and the fact that the next LTT for age 17 is in three years (2025), the value of an age 17 LTT PBA in 2022 might be questionable.

As discussed above, although not insurmountable, there are a few potential issues with the age 17 LTT in 2022. The budgetary aspects of this assessment will be discussed in closed session at the upcoming Governing Board meeting on May 14.