Committee on Standards, Design and Methodology

March 2, 2021

2:00 – 3:50 pm ET (Virtual)

AGENDA

2:00 – 2:05 pm	**Welcome and Overview of Agenda** *Gregory Cizek, Chair*	
2:05 – 2:15 pm	**Review and Revision of Mathematics and Reading Achievement Level Descriptions** *Sharyn Rosenberg, Asst. Director, Psychometrics*	Attachment A
2:15 – 2:25 pm	**Below the NAEP Basic Achievement Level** *Gregory Cizek*	Attachment B
2:25 – 2:35 pm	**Proposed Strategic Vision Activities** *Sharyn Rosenberg*	Attachment C
2:35 – 2:45 pm	**Framework Development Processes** *Gregory Cizek*	Attachment D
2:45 – 3:15 pm	**Next Steps for NAEP Linking Studies** *Gregory Cizek*	Attachment E
3:15 – 3:35 pm	**Status of the Achievement Levels Work Plan** *Gregory Cizek*	Attachment F
3:35 – 3:50 pm	**Next Steps** *Gregory Cizek*	

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 a total of 100 hours 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 (February 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 to panelists 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 consultation with the COSDAM Chair and Vice Chair, the Governing Board is planning to modify the project schedule to account for conducting the panel meetings in person in late 2021 and early 2022 An update on the project schedule and proposed design for convening in-person panel meetings will be presented during the May COSDAM meeting.

Below the NAEP Basic Achievement Level

One of the Governing Board's most important legislated responsibilities is developing the NAEP achievement levels. The Board policy on Developing Student Achievement Levels for NAEP defines three achievement levels: *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*. The policy specifies that, "The remaining region that falls below the *NAEP Basic* cut score shall be identified as 'below *NAEP Basic'* when a descriptor is necessary." The percentage of students in this lowest category is reported but there is no accompanying achievement level description (ALD).

During the December 2020 COSDAM meeting, there was discussion about the growing divergence of higher and lower performing students and the importance of better understanding what the lowest performing students know and can do. Some COSDAM members requested additional discussion about potential options such as describing the knowledge and skills that students may possess below the *NAEP Basic* level, and the merits and drawbacks of creating an ALD for below *NAEP Basic*.

Further discussion on these possibilities is tentatively planned for the May 2021 COSDAM meeting. To support that, two resources are currently under development to be included in the May 2021 COSDAM materials:

1. The Board has commissioned a paper to better understand:
    - the number of state assessments that have Below Basic ALDs;
    - the nature of Below Basic ALDs and how they differ from other categories;
    - the pros and cons of including Below Basic ALDs in state assessments; and
    - additional considerations relevant to NAEP.

2. The National Center for Education Statistics (NCES) convened an expert panel on December 16-17, 2020 to explore how NAEP can better measure, describe, and report the skills and knowledge of lower-performing students, particularly those below *NAEP Basic.* The panel consisted of the following members:

    George Bohrnstedt (Moderator), *American Institutes for Research*
    Henry Braun, *Boston College*
    Ray Hart, *Council of the Great City Schools*
    Hanseul Kang, *Yale University*
    Irwin Kirsch, *Educational Testing Service*
    Michele Mailhot, *Maine Department of Education*
    Pamela Mason, *Harvard University*
    Gary Phillips, *Cambrium Learning*
    Jennifer Randall, *University of Massachusetts, Amherst*
    Lorrie Shephard, *University of Colorado, Boulder*

    A report from the expert panel meeting is currently under development and will be disseminated to COSDAM when available.

Proposed Strategic Vision Activities

During a special Board meeting on September 29, the Board adopted its second strategic vision, Strategic Vision 2025, which is intended to guide the Board's work over the next four years. Strategic Vision 2025 includes three pillars (Inform, Innovate, and Engage) and eight priorities. Each strategic priority may include activities across several different standing committees, but it COSDAM will have primary responsibility for the following priorities:

Link NAEP resources with external data sources and disseminate what is learned from these sources so that NAEP can inform policy and practice in understandable and actionable ways

Develop a body of evidence to improve the interpretation and communication of NAEP achievement levels to ensure that they are reasonable, valid, and informative to the public

In addition, COSDAM may lead or be involved in activities that fall under any of the other strategic priorities.

Board staff have been developing proposed work plans for implementing each strategic priority.

During the upcoming COSDAM meeting, Committee members will briefly discuss proposed accomplishments for year 1. The discussions on NAEP linking studies and the Achievement Levels Work Plan at the upcoming meeting will also help inform potential next steps towards implementing these priorities.

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 is meeting its intended goals and determine whether any aspects need to be revisited.

To support a future discussion on this topic, Board staff have commissioned two papers that are currently under development:

- As a consultant, former Governing Board Executive Director Cornelia Orr is synthesizing 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) is preparing 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

Both papers are expected to be complete in the spring and will be shared in advance of future Board discussions.

Next Steps for Linking NAEP to External Data Sources

One of the goals of the Board's new Strategic Vision 2025 is to "Link NAEP resources with external data sources and disseminate what is learned from these sources so that NAEP can inform policy and practice in understandable and actionable ways." The Board's first Strategic Vision also included a goal of increasing opportunities to link NAEP results to external data, and the Achievement Levels Work Plan that the Board adopted last year includes a commitment to link to external data sources to provide additional context and relevance for NAEP results, including but not limited to the NAEP achievement levels. COSDAM has had several previous discussions about linking studies that have been completed, are currently underway, or could be undertaken with future administrations of NAEP.

The purpose of the March 2021 COSDAM session is to discuss potential next steps for developing a comprehensive strategy to address the Board's goals to link NAEP to external data sources as articulated by the Strategic Vision and the Achievement Levels Work Plan.

Overview

The NAEP assessment consists of cognitive items (the test questions) and contextual variables (the survey questions administered to students, teachers, and school administrators). The contextual variables are part of the NAEP administration and are intended to provide information to better understand the student achievement results (e.g., student experiences in and out of school, instruction, use of technology, available coursework).

The NAEP contextual variables are guided by the Board's Contextual Information Framework and the Board policy on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting. These policy documents provide guidance for collecting and reporting NAEP contextual variables as part of the NAEP assessment in accordance with the legislative requirement that in carrying out NAEP, the NCES Commissioner "only collect information that is directly related to the appraisal of academic achievement, and to the fair and accurate presentation of such information." A limited number of contextual variables are included with the results in the official release of the Nation's Report Card, with the majority available for secondary analyses through the NAEP Data Explorer and restricted-use datasets.

The term "external data sources" refers to additional data that may be useful in providing further context to understand NAEP results but is collected outside of the NAEP assessment and connected or linked to NAEP data through special studies. Data from external sources, including assessments and surveys outside of NAEP, sometimes can be connected to NAEP at the individual student level or at a higher level of aggregation such as the school or district level. There are no guiding documents for how and when to pursue studies that link NAEP to external data sources; existing studies have been initiated by NCES, NAEP contractors, and the Governing Board for a variety of purposes. For example, the studies that have been led and funded by the Governing Board have primarily been in support of the Board's research on academic preparedness for college.

Background on Previous NAEP Linking Studies

NAEP linking studies generally involve connecting data from a particular NAEP assessment to data from another assessment, providing information about where a NAEP score would fall on the scale of another assessment, and/or where a score from another assessment would fall on the NAEP scale. Scores from a NAEP assessment can be connected to scores from another assessment under the following conditions: 1) some items from another assessment are included as part of the administration of NAEP; 2) a common group of students takes both NAEP and another assessment (typically at different points in time); or 3) randomly equivalent groups of students take NAEP and another assessment. Linkages to other assessments are either concurrent (i.e., relating NAEP to another outcome that takes place within the same time frame) or predictive (i.e., relating NAEP to a future outcome).

In addition to referring to assessment data, the term "linking study" has also been used to describe efforts to connect information from NAEP to data from other NCES surveys. For example, the Early Childhood Longitudinal Study – Kindergarten (ECLS-K) and the High School Longitudinal Study (HSLS) include a parent questionnaire but NAEP does not; parent-reported data about income and occupation have been used in ongoing efforts to develop and validate socio-economic status (SES) measures for NAEP[1].

Previous NAEP linking studies have connected NAEP to other NCES surveys and longitudinal studies (via intra-agency agreements within NCES); data from state longitudinal databases (via agreements with state agencies); and external assessments (via agreements with other testing programs). Some of this work has been led by NCES while other studies have been initiated by the Governing Board and carried out in collaboration with NCES staff and contractors.

NAEP linking studies have been performed for a variety of purposes, such as:

- To estimate state-level performance on international assessments (e.g., linking NAEP and TIMSS was used to estimate TIMSS scores for all 50 states)
- To compare NAEP achievement levels with external benchmarks (and to understand the stringency of those performance standards) (e.g., a linking study of NAEP grade 4 reading and the Progress in International Reading Literacy Study [PIRLS]) found that the NAEP achievement levels are more stringent than the PIRLS benchmarks, explaining why fewer students reach the NAEP Proficient level in comparison to the PIRLS High benchmark)
- To compare state performance standards on a common scale (e.g., the state mapping studies use NAEP as a common metric for comparing the stringency of performance standards on state assessments)
- To estimate student performance on an external indicator of achievement, such as the percentage of grade 12 students academically prepared for college (e.g., several studies

[1] For several years, the NAEP program has been engaged in efforts to establish an improved measure of SES but no changes have been made to NAEP reporting of SES at this time; currently the program is still using eligibility for the National School Lunch Program and highest level of parental educational attainment.

including a national NAEP-SAT linking study were used to determine the point of the NAEP grade 12 reading and mathematics scales that corresponded to likelihood of placement in non-remedial college coursework)

- To inform the collection of non-academic outcomes, such as the development of a new measure of socio-economic status (e.g., a link between NAEP and parent questionnaires from ECLS-K and HSLS were used to explore new measures of SES for NAEP)

Several studies have been performed over the past 10-15 years or are currently underway. Attachment E2 briefly describes the purpose and conclusions of each study, with a link to completed reports (where available) for additional information. These existing studies generally focus on linking individual student data but it is also possible to connect NAEP data to external indicators at a higher level of aggregation such as the school, district, or state level. At a future meeting, NCES could provide more information on the feasibility of linking NAEP to institutional datasets.

Dissemination of NAEP Linking Studies

The existing NAEP linking studies are not available in a central location; some are official NCES reports and have been published on the NCES website, some are Governing Board reports and have been posted on the Governing Board website, some have been published by NAEP contractors and/or presented at conferences, and some are unpublished. The majority of the published reports are technical in nature (intended for researchers) and have not been adapted to materials appropriate for a more general audience.

Both the Strategic Vision and the Achievement Levels Work Plan include dissemination as an essential component of the goals related to NAEP linking studies, including synthesizing results from multiple studies and making the findings more accessible to non-technical audiences. To begin to explore potential ideas based on existing studies, a technical memo was commissioned last year under the Board's contract for Technical Support in Psychometrics, Assessment Development, and Preparedness for Postsecondary Endeavors. This memo and other strategies for reporting and disseminating findings from NAEP linking studies should be discussed at a future meeting in collaboration with the Reporting and Dissemination Committee.

Considerations for Pursuing Additional NAEP Linking Studies

Periodically, the Board provides input to NCES on the desirability of additional NAEP linking studies. Several factors affect the feasibility of undertaking new linking studies, including:

NAEP Assessment Schedule: Many linking studies are based on the same sample of students (or randomly equivalent groups of students) taking NAEP and another assessment. The Governing Board has taken this into account when making some decisions about the NAEP Assessment Schedule, such as ensuring that the NAEP assessments were administered in 2011 to link to the Trends in International Mathematics and Science Study (TIMSS) at grade 8 and to the Progress in International Reading Literacy Study (PIRLS) at grade 4. In some cases, the administration

years do not always align due to variation in periodicity (e.g., PIRLS is administered every 3 years and NAEP reading is administered every two years, so they only overlap once every 6 years). In addition, the recent shift of the NAEP reading and mathematics assessments from odd years to even years will have implications for alignment with other assessment schedules.

Legal requirements: In order to connect additional data to students in the NAEP sample, special permission is needed. This is the case whether the study involves another NCES data collection or an agreement with a state agency or external testing program. The data sharing agreements with state agencies and external testing programs typically have taken multiple years to negotiate with legal and contracts departments of multiple parties and have been very labor intensive. Many data sharing agreements are very specific in terms of what research questions can be addressed by the scope of the study; additional questions cannot be added later unless the agreements are re-negotiated. In addition, all NAEP research needs to comply with the NAEP legislation, including the prohibition against disclosing personally identifiable information; the linking procedures have included an elaborate process based on pseudo identification numbers.

Funding*:* Undertaking new linking studies is also a function of available funding. Some linking studies are much more expensive than others (e.g., when additional data collection is required compared to a naturally occurring overlap of samples). In some cases, the Governing Board has funded linking studies that have been performed at its request to support existing initiatives.

Content similarity*:* In order for results from a study linking NAEP to another assessment to be useful, a precursor step is to evaluate whether the constructs measured by NAEP and the other assessment are similar enough to allow for meaningful comparisons. Content alignment studies are generally performed in advance of conducting statistical linking studies.

Key Findings from NAEP Linking Studies

Over the past 5 years, COSDAM has had several discussions about studies that link NAEP to other surveys or assessments. NAEP linking studies have connected NAEP to other NCES surveys and longitudinal studies (via intra-agency agreements within NCES); data from state longitudinal databases (via agreements with state agencies); and external assessments (via agreements with other testing programs). Some of this work has been led by NCES while other studies have been initiated by the Governing Board and carried out in collaboration with NCES staff and contractors.

NAEP linking studies have been performed for a variety of purposes, such as:

- To estimate state-level performance on international assessments (e.g., linking NAEP and TIMSS was used to estimate TIMSS scores for all 50 states)
- To compare NAEP achievement levels with external benchmarks (and to understand the stringency of those performance standards) (e.g., a linking study of NAEP grade 4 reading and the Progress in International Reading Literacy Study [PIRLS]) found that the NAEP achievement levels are more stringent than the PIRLS benchmarks, explaining why fewer students reach the NAEP Proficient level in comparison to the PIRLS High benchmark)
- To compare state performance standards on a common scale (e.g., the state mapping studies use NAEP as a common metric for comparing the stringency of performance standards on state assessments)
- To estimate student performance on an external indicator of achievement, such as the percentage of grade 12 students academically prepared for college (e.g., several studies including a national NAEP-SAT linking study were used to determine the point of the NAEP grade 12 reading and mathematics scales that corresponded to likelihood of placement in non-remedial college coursework)
- To inform the collection of non-academic outcomes, such as the development of a new measure of socio-economic status (e.g., a link between NAEP and parent questionnaires from ECLS-K and HSLS were used to explore new measures of SES for NAEP)

Several studies have been performed over the past 10-15 years or are currently underway. Each study is summarized briefly below, including the design and methodology, key findings, and application to NAEP.

2007 NAEP-Early Childhood Longitudinal Study–Kindergarten (ECLS-K) Cohort of 1998-1999

Purpose: ECLS-K is a longitudinal study conducted by NCES to follow a cohort of students who entered kindergarten during the 1998-1999 school year through their eighth-grade year in 2006-2007. The study includes data collected from students, parents, teachers, and schools. The linking study served at least two purposes. One research study investigated the relationship between ECLS-K reading proficiency levels and 8th-grade NAEP achievement levels and explored the relationship between reading performance at earlier grades and performance on the 8th-grade NAEP reading assessment. Another research study investigated the concordance of student-reported parental education on the NAEP student background questionnaire with parent reports on the same variable from the ECLS-K questionnaire.

Sample: Data came from a common sample of public school students (n=1,290) who took both NAEP and ECLS-K grade 8 reading assessments in spring of 2007.

Statistical method to establish the link: Projection by regression was used in this study.

Main findings: The correlation between NAEP Reading and ECLS Reading at grade 8 was estimated at r = .83.

Reading Analysis: The link allowed a comparison between NAEP grade 8 achievement levels in reading and the finer grain and developmentally descriptive ECLS reading proficiency levels. Reading skills students need to master in earlier grades to later reach NAEP's *Proficient* level at grade 8 were identified.

> Dogan, E., Ogut, B., & Kim, Y. (2015). Early childhood reading skills and proficiency in NAEP eighth-grade reading assessment. *Applied Measurement in Education, 28(3),* 187-201.

Parental Education Analysis: With few exceptions, the higher the parent's education, the more accurate the student estimates are of what their parent's education is as reported by one of the parents. Consistent with this result, the higher the parent's education, the lower the percentage of students who report "I don't know". The high polychoric correlations computed with the "don't knows" eliminated and the relatively small bias in analyses using student-reported parental education instead of parent-reported suggest that in spite of the inaccuracies in student reports of parental education, valuable information is nonetheless contained in students' reports of parental education.

> Ogut, B. and Bohrnstedt, G. W. (2012). Reliability of student-reported parental education at NAEP grade 8 mathematics assessment. Paper presented at the annual meeting of the American Educational Research Association, Vancouver.

Application to NAEP: Information from this study on SES is being considered among other pieces of information in the formulation of a new SES measure.

2009 Preparedness Research: Statistical Linking of NAEP and the SAT

Purpose: This study was conducted as part of the Governing Board's research program on using NAEP as an indicator of academic preparedness for college. The purpose of this study was to identify a reference point or range on the NAEP 12th grade reading and mathematics scales that might be associated with the College Board's SAT preparedness benchmarks. The NAEP and SAT scores for 12th grade students who had taken NAEP in 2009 and had also taken the SAT were the basis for this linking (via an agreement with the College Board).

Sample: The overall NAEP sample size for 2009 12th grade was 49,000 (reading) and 46,000 (math). Students who also took the SAT were matched to NAEP resulting in 16,200 students (reading) and 15,300 students (mathematics), or approximately 33% of students. Note this was conducted for public-school students only. This match rate compares favorably to the national SAT participation rate of approximately 36% of public school students.

Statistical method used to establish the link: Two types of statistical linking were considered in this study: concordance and projection. Projection was preferred primarily due to the moderate correlation of 0.74 for NAEP reading and SAT-reading. (The correlation for math was 0.91.)

Main findings: Based on the College Board's designation of 500 as the preparedness benchmark for each subject at the time the study was conducted, using statistical projection defined the preparedness estimates for NAEP at 302 (reading) and 164 (math). Note that 302 is the reading NAEP Proficient cut score and 176 is the math NAEP Proficient cut score. A report of the results is available on the Governing Board website at (link to NAEP/SAT Report).

Application to NAEP: Findings from this study and others were used to report estimates of the percentage of students academically prepared for college in the 2013 and 2015 NAEP grade 12 report cards. A similar methodology was applied to a linking study of 2013 12th grade NAEP and ACT data at the national level (via a data sharing agreement with ACT) and for a few states (via data sharing agreements with states). In addition, 2013 12th grade NAEP and SAT scores were linked for students in one state via a data sharing agreement with Massachusetts.

2009 Preparedness Research: Longitudinal Analyses of Performance on NAEP Related to Performance in College and Other Outcomes of Florida Students:

Purpose: This study was conducted as part of the Governing Board's research program on using NAEP as an indicator of academic preparedness for college. The purpose of this study was to relate 2009 grade 12 NAEP scores to ACT and SAT scores, college performance and other outcomes. Working with Florida state officials and their K-20 Education Data Warehouse (a longitudinal database) scores for students who had participated in the 2009 NAEP 12th-grade assessments and were subsequently enrolled in Florida's public colleges in 2010 were linked to a variety of outcome indicators.

Sample: The overall NAEP sample size for 2009 Florida 12th grade was 3,400 (reading) and 3,200 (math). Sample size for students attending Florida public colleges in 2010 was 1,800 (math) and 1,900 (reading), or about 55% of the NAEP-sampled students. Approximately one-third of these students attended 4-year colleges and about two-thirds attended community colleges.

Statistical method: Average 2009 grade 12 NAEP scores (and interquartile ranges) were reported for seven variables related to postsecondary performance: SAT preparedness benchmarks; ACT preparedness benchmarks; Accuplacer performance; students' self-reported program of study in high school; college enrollment; first year college coursetaking; and first year grade point average.

Main findings: Based on the College Board's designation of 500 as the preparedness benchmark for each subject, 53% of Florida's 12th graders were deemed college ready for mathematics and 54% were for critical reading. Based on the ACT benchmarks of 22 for mathematics and 21 for reading, 34% of Florida's 12th graders were college-ready for mathematics and 46% were college-ready in critical reading. Finally, first year of college results showed a greater percentage of students achieving GPA of B- or better during their first year of college scored at or above the potential NAEP preparedness reference points from the NAEP-SAT linking study compared to students whose GPA was less than a B- during their first year of college. The limitations of the Florida data, namely the availability of data only for students enrolled in Florida public postsecondary institutions, must be taken into consideration when interpreting these results. The report can be found on the Governing Board website: (link to Florida report).

Application to NAEP: Findings from this study and others were used to report estimates of the percentage of students academically prepared for college beginning with the 2013 NAEP grade 12 report card. Longitudinal research includes a few additional state partners for 2013 NAEP.

2011 NAEP-Trends In Mathematics and Science Study (TIMSS) Linking Study

Purpose: TIMSS is an international comparison study of student achievement in mathematics and science at grades 4 and 8, administered every four years. The purpose of conducting the 2011 NAEP-TIMSS linking study was two-fold. The study was conducted to see whether it is possible to predict TIMSS scores (in mathematics and science) for the states that did not participate in the TIMSS assessment. Secondly, the study was conducted to identify a method among various methodologies suggested in the literature for linking two assessments. The study was done at grade 8 only.

Sample: The study involved four samples of students at grade 8: the 2011 NAEP operational/national sample, the 2011 TIMSS U.S. operational/national sample, students assessed using 2011 NAEP administration procedures who received braided booklets containing one block of NAEP and one block of TIMSS items; and students assessed using 2011 TIMSS administration procedures who received one block of NAEP items and three blocks of TIMSS items. In addition to these linking study samples, nine states—Alabama, California, Colorado, Connecticut, Indiana, Florida, Massachusetts, Minnesota, and North Carolina—participated in 2011 TIMSS as separate jurisdictions to serve as the "validation sample".

Statistical method used to establish the link: Three types of statistical linking were considered in this study: statistical moderation, statistical projection, and IRT calibration.

Main findings: Selected findings are highlighted below (link to NAEP-TIMSS linking study report).

For Mathematics:

- Average scores for public school students in 36 states were higher than the TIMSS average of 500.
- Scores ranged from 466 for Alabama to 561 for Massachusetts.
- Massachusetts scored higher than 42 of the 47 participating education systems.
- Alabama scored higher than 19 education systems.

For Science:

- Average scores for public school students in 47 states were higher than the TIMSS average of 500.
- Scores ranged from 453 for the District of Columbia to 567 for Massachusetts.
- Massachusetts and Vermont scored higher than 43 participating education systems.
- The District of Columbia scored higher than 14 education systems.

The evaluation of results showed that all three methods of linking yielded essentially the same predicted TIMSS results. In addition, among the three methods, the statistical moderation technique is the simplest method requiring the estimation of the fewest parameters and could be applied to the extant national samples of NAEP and TIMSS. (link to NAEP-TIMSS linking study technical report).

Application to NAEP: The predicted TIMSS scores for states were reported and compared to other countries. This study also helps NCES conduct future NAEP-TIMSS linking studies using statistical moderation without the additional resources needed for the braided-booklet samples.

2011 NAEP- Progress in International Reading Literacy Study (PIRLS) Linking Study

Purpose: PIRLS is an international comparison study of reading literacy at grade 4, administered every five years. The purpose of this study was to obtain a statistical comparison between NAEP and PIRLS. The results of the 2011 NAEP grade 4 reading assessment were expressed in terms of the metric of the 2011 PIRLS assessment thereby providing international benchmarks for the NAEP grade 4 reading achievement levels.

Sample: Separate operational national samples of 2011 NAEP and 2011 PIRLS (the design did not include administering both assessments to a common sample of students). Florida did participate in 2011 PIRLS at the state level and was used to validate the linking results.

Statistical method to establish the link: Statistical moderation was used.

Main findings: At each level, the linking shows that the NAEP grade 4 reading achievement levels are higher than the PIRLS international benchmarks. The study report can be found at: http://files.eric.ed.gov/fulltext/ED545246.pdf

When the actual PIRLS results for Florida were compared to the projected PIRLS results, the mean difference was not statistically significant. The only significant difference between the two sets of results for Florida was for the percentage of Advanced students (which varied by only one percentage point).

Application to NAEP: The fact that NAEP reading achievement levels are higher than similar PIRLS international benchmarks may help explain why NAEP has historically reported lower rates of reading proficiency for the United States, whereas PIRLS has historically reported higher levels of reading proficiency. For example, in 2011, NAEP reported that 34 percent of fourth graders were reading at the proficient level, while PIRLS reported that 56 percent were reading at the high international benchmark.

2013 NAEP-PISA Linking Study

Purpose: The pilot 2013 NAEP-PISA linking study had three goals:
1) Examine the possibility of creating a NAEP mathematics cross-grade scale for the purposes of obtaining a NAEP mathematics score distribution for 15-year-olds, given that NAEP typically does not assess 15-year-old students.

2) Examine the difference in student performance between grades 8, 9, 10, 11, and 12 on the NAEP mathematics scale, based on the cross-grade scale developed under goal 1), and investigate whether statements about annual growth can be made.

3) Examine the possibility of creating a horizontal link between the NAEP cross-grade mathematics scale and PISA mathematics literacy scale using statistical moderation under an assumption of randomly equivalent samples (15–year-olds).

The study was a way to pilot the creation of a direct statistical link between grade-based NAEP and age-based PISA, given significant differences between the two assessments. In other words, a future linking study could include a national component in NAEP that has off-grade administrations to establish a more defensible link and this is the first step towards that. If a cross-grade scale can be established successfully and a link with PISA is defensible, the ultimate goal is to estimate PISA scores for U.S. states who participate in NAEP.

Sample: Two states that participated in the 2013 NAEP state-level 12th-grade pilot and had participated in the 2012 PISA were included in this study. In each state, additional samples of students in grades 9, 10, and 11 were administered a version of the NAEP mathematics assessment.

Statistical Method: Two states, Massachusetts and Florida, had previously (2012) participated in PISA, and students in each state were selected to participate in this special study. The students were selected from grades 8, 9, 10, 11, and 12.

While new NAEP mathematics assessment content was not developed, test booklets administered at grades 9, 10 and 11 were specially configured with NAEP operational grade 8 and/or grade 12 blocks. Three types of booklets were administered in the study to students according to the scheme in table below:

- Grade 8-only NAEP operational mathematics books
- Grade 12-only NAEP operational mathematics books
- Braided books comprised of one grade 8 NAEP operational mathematics block and one grade 12 NAEP operational mathematics block.

Targeted Percentages of Students Assessed, by Book Type

	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12
Grade 8-only books	100%	100%	25%	15%	0%
Braided 8/12 books	0%	0%	50%	50%	0%
Grade 12-only books	0%	0%	25%	35%	100%
Effective 8th grade content exposure rate	100%	100%	50%	40%	0%
Effective 12th grade content exposure rate	0%	0%	50%	60%	100%

Results: Although it was determined that establishing a statistical link between NAEP and PISA may be feasible through linear equating, the validity of the predicted PISA results requires further evaluation. Some questions remain to be addressed in terms of the validity of the linking results, such as the constructs measured by NAEP and PISA, the definition of the target populations between NAEP and PISA given differences in the timing window (different assessment years), exclusion policies, etc. Also, the NAEP-PISA link was established based on this work.

Application to NAEP: A NAEP-PISA linkage might be sought in the future to provide a basis for international comparisons.

2013 NAEP-High School Longitudinal Study (HSLS)

Purpose: HSLS is a longitudinal study conducted by NCES to follow a cohort of students who were in ninth grade during the 2009-2010 school year throughout their secondary years and into their postsecondary years. Data for students who had participated in both the 2013 NAEP 12th grade assessments and the HSLS were linked so that information from the HSLS student and parent questionnaires could provide a broader context for understanding NAEP results. In addition, the study explored using the relationship between the HSLS questionnaire variables and NAEP scores to predict NAEP mathematics scale scores for the full HSLS sample. The results from this research study are under review by NCES.

Sample: Students in the HSLS study who were also tested in NAEP in the 12th grade. N = 3,471 NAEP 2013 Math; 717 NAEP 2013 Reading.

Statistical method to establish the link: Imputation by multiple regression.

Main findings: The results from regression analyses and validation tests show that it is feasible to impute NAEP scale scores with acceptable accuracy for the full ~20,000 HSLS sample using data from the NAEP-HSLS overlap sample (N=3,471). Specifically, models that use HSLS algebra performance in grades 11 and 9 combined with student student-level covariates including race/ethnicity, gender, SD status, ELL status, and parental education proved to work best in recovering actual mean scores of student subgroups from the HSLS-NAEP overlap sample. The pseudo-R-squared of the best fitting model with the least bias was 0.744 (R = .863).

Additional analysis examined how Grade 12 NAEP mathematics achievement predicts enrollment in postsecondary education with or without remediation, selectivity of the colleges enrolled, persistence in postsecondary education, and majoring in a STEM field. The second set of analyses examined the ability of Grade 12 NAEP mathematics achievement in predicting postsecondary outcomes relative to SAT-Mathematics achievement, with and without controlling for high school GPA.

Application to NAEP: There are multiple applications. For example, the study that investigated SES in the NAEP overlap sample and follow-on research resulting from this study (as well as additional similar efforts proposed for the NAEP-ECLS-K overlap sample of 2015) could inform the development of a simple and effective SES index based on student level SES items (existing one and/or newly piloted ones). Also possible with the HSLS is the derivation of preparedness benchmarks for college attendance and graduation, derivation of regression models predicting students' entrance into college stem fields, and validation of noncognitive measures as correlates of NAEP Mathematics and predictors of postsecondary outcomes.

2013 NAEP-EXPLORE (KY, NC, TN)

Purpose: The ACT Explore assessments were designed to assess a specific student's academic progress at the 8th or 9th grade levels, especially with respect to college and career readiness. As part of the Governing Board's research on using NAEP to estimate the percentage of students academically prepared for college, the NAEP-EXPLORE linking studies tried to identify reasonable points on the grade 8 NAEP reading and mathematics scales that indicate being on track for academic preparedness for college by the end of high school.

Sample:

- 3,700 and 3,800 for reading and math respectively in KY (including TUDA sample), and overall matching rates are 96% for both subjects.
- 4,000 and 3,900 for reading and math respectively in NC (including TUDA sample), and overall matching rates are 96% for both subjects.
- 2,700 each for reading and math in TN, and overall matching rates are 93% and 94% respectively.

Statistical method: Given that the correlation between NAEP and EXPLORE was not strong enough to support concordance, it was decided a statistical projection was a more appropriate choice. The correlations ranged from 0.72 to 0.74 for reading and from 0.81 to 0.82 for mathematics.

Main findings: In general, the relationship between NAEP and EXPLORE is moderate. Based on the Explore benchmarks of 16 for reading and 17 for mathematics, the NAEP *Proficient* achievement levels for reading and mathematics at grade 8 correspond well with the EXPLORE benchmarks and could possibly be used to form reasonable basis for reporting 'on track for preparedness'. The reports can be found on the Governing Board website: KY, NC, and TN.

Application to NAEP: Results have not been applied to operational NAEP but could potentially be used to explore the feasibility of reporting estimates of the percentage of students on track to be academically prepared for college by the end of high school.

2013 NAEP-SAT Linking Study in MA – Grade 12

Purpose: As the second phase of the Governing Board's academic preparedness research, state-level statistical linking studies were conducted between NAEP and either SAT or ACT to identify a reference point on the NAEP 12th grade reading and mathematics scales that might be associated with the existing college readiness benchmarks established by College Board (for SAT) or ACT (for ACT). In this study, the NAEP and SAT scores for the 12th graders in Massachusetts who had taken NAEP in 2013 and had also taken the SAT were the basis for this linking (via an agreement with the Massachusetts Department of Education).

Sample: 2,400 MA students for reading with an overall matching rate of 74%, and 2,400 MA students for mathematics with an overall matching rate of 76%.

Statistical method used to establish the link: Given the different assessment purposes of NAEP and SAT and the moderate correlation between NAEP reading and SAT critical reading (r=0.74), it was decided a statistical projection, instead of concordance, was a more appropriate choice for linking the two scales on reading and mathematics. The correlations for mathematics was 0.89.

Main findings: The SAT benchmarks (i.e., 500 for each subject) and the NAEP *Proficient* achievement level cut scores correspond well to each other for reading in both linking directions, but somewhat differ for mathematics. In particular, the NAEP reading *Proficient* achievement level cut score of 302 could form a reasonable basis for reporting on academic preparedness for college at grade 12 in Massachusetts, while the mathematics counterpart is 164 on the NAEP scale, about 12 points lower than the NAEP *Proficient* achievement level cut score for grade 12 mathematics. These two cut scores identified using the 2013 Massachusetts NAEP-SAT linking sample coincide with the corresponding cut scores defined by the 2009 NAEP-SAT linking study. A report of the results is available on the Governing Board website (link to the Massachusetts NAEP-SAT linking study report).

Application to NAEP: Results suggested that the statistical relationship between NAEP and SAT established for the Massachusetts linking sample surveyed in the 2013 NAEP assessment is very similar to that established with the 2009 NAEP-SAT linking samples on the national level, which could be used to support the validity of the preparedness benchmarks identified in the 2009 NAEP-SAT linking study.

2013 NAEP-ACT Linking Studies in TN and MI – Grade 12

Purpose: As the second phase of the Governing Board's academic preparedness research, state-level statistical linking studies were conducted between NAEP and either SAT or ACT to identify a reference point on the NAEP 12th grade reading and mathematics scales that might be associated with the existing college readiness benchmarks established by College Board (for SAT) or ACT (for ACT). In these two studies, the NAEP and ACT scores for the 12th graders in each state who had taken NAEP in 2013 and had also taken the ACT were the basis for the linking (via agreements with the Tennessee Department of Education and the Michigan Department of Education).

Sample:

- 3,000 and 3,200 for reading and mathematics respectively in TN, and the overall matching rates were 89% and 90% for reading and mathematics respectively.
- 2,900 and 3,100 for reading and mathematics respectively in MI, and overall matching rates were 95% for both subjects.

Statistical method used to establish the link: Given that the correlation between NAEP and ACT was not strong enough to support concordance, it was decided a statistical projection was a more appropriate choice. The correlation was 0.73 for reading and 0.83 for mathematics, for both states.

Main findings: In general, the relationship between NAEP and ACT was moderate, based on the linking samples from Tennessee and Michigan. Results showed that the ACT college readiness benchmarks and the NAEP *Proficient* achievement level cut scores correspond well to each other for reading in both linking directions but differ more for mathematics. (The NAEP reading *Proficient* and the NAEP math *Proficient* achievement level cut scores are 302, and 176, respectively.) In the state of Tennessee, the NAEP reading scale score of 301 could form a reasonable basis for reporting on academic preparedness for college, while the mathematics counterpart is 168 on the NAEP scale. In the state of Michigan, the NAEP reading scale score of 308 could form a reasonable basis for reporting on academic preparedness for college, while the mathematics counterpart is 169 on the NAEP scale. Reports of the results is available on the Governing Board website (link to the Tennessee NAEP-ACT linking study report and link to the Michigan NAEP-ACT linking study report).

Application to NAEP: Results have not been applied to operational NAEP.

2013 NAEP-ACT National Linking Study – Grade 12

Purpose: This study was conducted as part of the second phase of the Governing Board's academic preparedness research. The purpose of this study was to identify a reference point on the NAEP 12th grade reading and mathematics scales that might be associated with the ACT's college readiness benchmarks. The NAEP and ACT scores for 12th grade students who had taken NAEP in 2013 and had also taken the ACT were the basis for this linking (via an agreement with ACT).

Sample: Overall, approximately 44,300 twelfth-grade public school students were assessed in NAEP reading and 44,900 twelfth-grade public school students were assessed in NAEP mathematics in 2013. NAEP scores were matched at a rate of 41% for the weighted reading sample and 42% for the weighted mathematics sample, resulting in 19,900 students for reading and 20,300 students for mathematics. These match rates were lower than the national ACT participation rate of approximately 54% of high school graduates[1] in 2013.

Statistical method: Given that the correlation between NAEP and ACT was not strong enough to support concordance, it was decided a statistical projection was a more appropriate choice. The correlation was 0.75 for reading and 0.87 for mathematics.

Main findings: Based on the 2013 national sample, the relationship between NAEP and ACT was moderate. The results showed that the ACT college readiness benchmarks and the *NAEP Proficient* achievement level cut scores correspond well to each other for reading in both linking directions, but they differed slightly for mathematics. In particular, the reading *NAEP Proficient* achievement level cut score of 302 could form a reasonable basis for reporting on meeting the ACT college readiness benchmarks and therefore potentially academic preparedness for college at grade 12. However, the mathematics counterpart is 167 on the NAEP scale, about 9 points lower than the *NAEP Proficient* achievement level cut score for grade 12 mathematics. A report of the results is available on the Governing Board website (link to the NAEP-ACT linking study report).

Application to NAEP: Results from the 2013 NAEP-ACT national linking study suggested that the reading *NAEP Proficient* achievement level cut score of 302 could form a reasonable basis for reporting college readiness, confirming a similar conclusion from the 2009 NAEP-SAT national linking study.

[1] The NAEP-ACT linking sample was based on students attending public schools. The ACT participation rate, however, did not differentiate public school students from private school students. It was therefore anticipated that the match rate for the linking study would be lower than the ACT participation rate.

Longitudinal Analyses of Performance on 2013 NAEP Related to Performance in College of Michigan Students

Purpose: The purpose of this study was to relate NAEP scores to college performance for the Michigan public school students who participated in the 2013 grade 12 NAEP reading or mathematics assessment. The Michigan Department of Education provided six years of longitudinal data that cover year 1 through year 6 out of high school for those 12th graders who took grade 12 NAEP assessments in 2013, including college enrollment, remedial course taking, GPA, and degrees obtained. The NAEP data matched to the Michigan longitudinal data served as the basis on which the relationship between NAEP and college performance was derived.

Sample: In 2013, about 2,900 and 3,100 12th graders in Michigan were assessed for the NAEP reading and mathematics respectively, of which 98% were matched to the longitudinal data provided by the Michigan Department of Education.

Statistical method: Descriptive statistics such as mean, 25th, and 75th percentile of the NAEP scale score were computed for groups of students categorized by their college performance such as whether they had taken any remedial courses, whether they had a first year GPA of B- or above, etc. A comparison was then made on the average NAEP reading and mathematics scores (and IQR[2]) between those who were associated with a higher college performance indicator to those who were associated with a lower college performance indicator.

Main findings: Results suggested that for the matched Michigan 12th graders who took NAEP in 2013, their grade 12 NAEP performance is positively related to their performance in college evaluated in this report. For instance, students who enrolled in college and never took remedial courses are associated with better performance on the grade 12 NAEP Reading and Math assessments compared to those who had taken at least one remedial course in college. A report of the results is available on the Governing Board website: link to longitudinal MI report.

Application to NAEP: The average grade 12 NAEP reading and mathematics scale scores of those who never took remedial courses were very close to the preparedness benchmark identified for each subject by the 2009 NAEP-SAT linking study, which could be seen as confirmative and validity evidence for using the preparedness benchmarks to indicate 12th graders' academic preparedness for college.

[2] IQR: inter-quartile range, the difference between the 75th and 25th percentiles

Longitudinal Analyses of Performance on 2013 NAEP Related to Performance in College of Massachusetts Students

Purpose: The purpose of this study is to relate NAEP scores to college performance for the Massachusetts public school students who participated in the 2013 grade 12 NAEP reading or mathematics assessment. The Massachusetts Department of Education agreed to provide six years of longitudinal data that cover year 1 through year 6 out of high school for those 12th graders who took grade 12 NAEP assessments in 2013, including college enrollment, remedial course taking, GPA, and degrees obtained. The NAEP data matched to the Massachusetts longitudinal data will serve as the basis for the relationship between NAEP and college performance.

Analyses are intended to be similar to the previous description of the longitudinal study in Michigan, but not all of the data have been received yet to conduct these analyses.

2013 NAEP-Lexile® Study

Purpose: The Lexile® framework and measures (owned by MetaMetrics®) include a vertical reading scale that spans grades 1 to 12, in addition to benchmarks for college and career readiness. The purpose of the study was to identify scores on the NAEP scale that correspond to preparedness benchmarks on the Lexile® scale. To accomplish this link, a subsample of students in the 2013 NAEP assessment were administered Lexile items. The NAEP-Lexile® study was successful in demonstrating a strong relationship between NAEP reading and the Lexile® measure of comprehension (r = .89). At the recommendation of the NAEP Design and Analysis Committee, ETS attempted to further evaluate the validity of the linking by comparing the estimated proportions of readiness based on the state-level linking results with the proportions estimated by this linking study. Data from only two states were available and several more would be needed to evaluate consistencies among the results. So, at this stage, the validity or generalizability of the study results are still inconclusive. No report has been published on this work.

Sample: A fraction of the total 12th-grade NAEP sample in Reading was used in this special study. Two groups were created by random assignment: 5,000 students responded to "braided" booklets consisting of one block NAEP and one block Lexile®; and 3,000 students responded to straight Lexile® booklets.

Statistical Method: Based on the study design, the NAEP Design and Analysis Committee (DAC) recommended using both statistical moderation and statistical projection approaches to establish the linkage between the two instruments.

Results: The NAEP-Lexile® study was successful in demonstrating a strong relationship between NAEP reading and the Lexile measure of comprehension (r = .89). At the recommendation of the NAEP Design and Analysis Committee, ETS attempted to further evaluate the validity of the linking by comparing the estimated proportions of readiness based on the state-level linking results with the proportions estimated by this linking study. Data from only two states were available and several more would be needed to evaluate consistencies among the results. So, at this stage, the validity or generalizability of the study results are still inconclusive. No report has been published on this work.

Applications to NAEP: The Lexile® scale would provide a means for estimating the percentage of students who meet or exceed the MetaMetrics® benchmarks for college and career readiness.

2015 NAEP-ECLS Kindergarten Cohort of 2010-2011

Purpose: ECLS-K is a longitudinal study conducted by NCES to follow a cohort of students who entered kindergarten during the 2010-2011 school year through their fifth-grade year in 2015-2016. The study includes data collection from students, parents, teachers, schools, and care providers. The parent interviews include information about income and parental education. The aim of the NAEP/ECLS-K special study is to evaluate the accuracy of grade 4 student reported parental occupation and education (the piloted NAEP SES-related questions), using the ECLS-K parent reported occupation and education as a reference. The results will be useful to inform development and interpretation of SES measures.

Sample: About 1,500 grade 4 students were assessed for both NAEP and ECLS-K in 2015 and were given an extended NAEP student questionnaire. The extended student questionnaire included a set of SES questions on parental occupation and education which are also being administered as part of the 2016 NAEP pilots and were tested in cognitive interviews prior to administration in the special study.

Statistical method used to establish the link: Data from the ECLS-K and NAEP datasets were merged by matching students based on common identification. Where available, one or both parents were interviewed as part of the 2015 ECLS-K grade 4 data collection, including SES-related questions of occupation and education. For households with two parents, the mother and father were interviewed separately.

Main findings: Analysis has been completed. Grade 4 students were unable to accurately report on the highest educational attainment of their parents and were unable to accurately report parental occupation.

Application to NAEP: The primary goal of this study was to evaluate the accuracy of grade 4 student reported parental occupation, parental education, and household income related questions, using the ECLS-K parent-reported occupation, education, and household income as a reference, in order to inform plans for collection of this information in NAEP, which could potentially be included as part of an SES measure. An internal report by ETS has been submitted. A secondary goal was to explore the growth of reading skills as a student progresses from kindergarten to fourth grade. This analysis has been completed and a formal report is being prepared.

2017 NCES Mapping of State Proficiency Standards

Purpose: NCES has periodically published reports on comparisons of the reading and mathematics proficiency standards states set for their public school students at grades 4 and 8. Performance standards vary across states, the results of the various state assessments cannot be used to directly compare students' progress. However, by placing or "mapping" a state standard onto the NAEP scale, a common metric for all states, a NAEP equivalent score of that standard is produced, which can be compared across states. Since 2003, NCES has conducted this study known as "state mapping" study eight times. This study helps examine (a) how a state's standard of proficient performance compare with those of other states, (b) where a state's standard maps on the NAEP scale in relation to the cut-scores of NAEP achievement levels.

Data sources: The study requires two sets of data for each of 50 states, District of Columbia, and Puerto Rico be included in the study:

a) State assessment results, that is, percentages of students performed at or above the respective state's standard for proficient performance in all public schools that participated in the NAEP reading and mathematics assessments in grades 4 and 8. NCES obtains the state data from a file developed for the Elementary and Secondary Education, a department of the federal government;
b) NAEP data for all public schools representing the states. Puerto Rico is included for mapping their proficiency standards in mathematics.

Statistical method to establish the link: By comparing the percentages of students in each NAEP school who achieve each of a state's performance standards with the distribution of NAEP performance by the random sample of students participating in NAEP in the school, we can approximately estimate the point on a common scale for the state standards. The method employed to map the state standards on the NAEP scale is known as equipercentile equating. Detailed information on the estimation methods is available at http://nces.ed.gov/nationsreportcard/pdf/studies/2010456.pdf.

Main findings: Results discussed here are from the latest mapping study available to the public, which was conducted using the NAEP 2017 assessment data and state assessment data for the public school for school year 2016-17.

Overall, in 2017, most state standards for both grades and both subjects mapped at the *NAEP Basic* achievement level. In addition, for states with three years (2017, 2015, 2007) of data, the difference between the highest and lowest NAEP equivalent scores of the state standards was smaller in 2017 than in 2015 and 2007 for each grade and subject, with the exception of grade 8 mathematics standards. More specifically, in 2017

In grade 4 reading:

a. Two of the 50 states included in the study had standards at the *NAEP Proficient* level, while three states had standards below the *NAEP Basic* level; all three testing programs—ACT, PARCC, and SBAC—had standards that mapped at the *NAEP Basic* level.
b. The difference between the highest and lowest NAEP equivalent scores was 34 points; this difference is larger than the difference between the grade 4 reading cut scores for the *NAEP Basic* and *NAEP Proficient* levels.

In grade 4 mathematics:

a. Eleven of the 51 states included in the study had standards at the *NAEP Proficient* level; the PARCC standard mapped at the *NAEP Proficient* level. ACT and SBAC standards mapped at the NAEP Basic level.
b. The difference between the highest and lowest NAEP equivalent scores was 33 points, which is smaller than the difference between the grade 4 mathematics cut scores for the *NAEP Basic* and *NAEP Proficient* levels.

In grade 8 reading:

a. Five of the 48 states included in the study had standards at the *NAEP Proficient* level; all three testing programs—ACT, PARCC, and SBAC—had standards that mapped at the *NAEP Basic* level.
b. The difference between the highest and lowest NAEP equivalent scores was 45 points, which is larger than the difference between the grade 8 reading cut scores for the *NAEP Basic* and *NAEP Proficient* levels

In grade 8 mathematics:

a. Seven of the 32 states included in the study had standards at the *NAEP Proficient* level; the ACT and SBAC standards mapped at the *NAEP Basic* level. The PARCC standard was not estimated because the states participating in PARCC did not require all grade 8 students to take a general mathematics assessment.
b. The difference between the highest and lowest NAEP equivalent scores was 33 points, which is smaller than the difference between the grade 8 mathematics cut scores for the *NAEP Basic* and *NAEP Proficient* levels.

It should be noted that not all states were included in each combination of grades and subjects for various reasons. Details are available in the report https://nces.ed.gov/nationsreportcard/subject/publications/studies/pdf/2019040.pdf

Application to NAEP: The NCES mapping study has contributed to the discussion on achievement standards for the nation's students since 2003. Certain cautions in interpretation are urged. The study is not an evaluation of the various state assessments or of the quality of the states' achievement standards, and the findings should not be interpreted as evidence of deficiencies in state assessments or in NAEP. It should be noted that state assessments and NAEP may vary in format and administration because they have different goals and are developed for different purposes.

Study of the Alignment of 2017 NAEP with Current Generation State Assessments Based on College and Career Ready (CCR) Standards: Quantitative Analysis to Support Item Mapping

Purpose: This research mapped items from three statewide assessments, along with NAEP mathematics and reading items, onto the NAEP scale. The purpose of this exercise was to complement reports by Daro and Hughes (2019) and Valencia et al. (2019) which used expert judgments of items in mathematics and ELA from four statewide assessment programs and NAEP in grades 4 and 8. Expert panels used structured protocols to provide qualitative and quantitative descriptions of the content covered, and the ways that content is covered, by the statewide assessments and NAEP, with an eye toward the degree to which NAEP measures similar aspects of achievement. Three of the four states whose assessments were studied by Daro and Hughes (2019) and Valencia et al. (2019) (denoted in these studies SA1, SA2, and SA3) also provided student and item level data, which was used in this study to link those states' assessment scales to the NAEP scale, and then to map the states' items onto the NAEP scale, interspersed with NAEP's items. The purpose was to illustrate the relationship (if any) between the item attributes, such as content, focus, and complexity as cataloged by Daro and Hughes (2019) and Valencia et al. (2019), with empirical item difficulty, and to compare the distribution of item difficulty on NAEP to that on state assessments. The three states studied here include one using an assessment from one of the multistate consortia, and two states that use their own assessments but that also reflect College and Career Readiness (CCR) standards to varying degrees.

Sample:
For some analyses, all students in the state were used, but for the core linking analysis, the overlap sample of students who took both the state assessment and NAEP was used.

Statistical method:
The data used for this study included item- and student-level information for NAEP and state assessments from three states, one of which used a consortium test while the other two used their own state tests. The NAEP item-level data included the IRT parameters for all items in the 2017 operational Mathematics and Reading assessments at Grades 4 and 8, and identifiers that associated the items with the expert judgment study ratings. The NAEP student-level data included student responses to cognitive items as well as background demographics. Similarly, the study used IRT parameters for all operational items in the state Math and ELA assessments at Grades 4 and 8 to conduct the statistical linking. From each of the three states, the study collected student-level data for all students in Grades 4 and 8 including responses to each assessment item, the total raw score, and sub-scale scores.

The states' mathematics assessments were linked to NAEP mathematics, and the subset of the states' ELA assessment items most closely related to reading was linked to NAEP reading. Then standard NAEP item mapping protocols (Donoghue, 1997) were used to map both the states' and NAEP items onto the NAEP scale. Those item maps were then combined with information from the studies by Daro and Hughes (2019) and Valencia et al. (2019) to illustrate the relations between item attributes and the degree of empirical difficulty of the items as reflected by their placement on the item maps.

Main findings: The state assessments correlated strongly with NAEP. At the composite level the correlations between NAEP and the state assessments ranged from .88 to .98, with correlations for reading ranging from .88 to .94 and for mathematics ranging from .93 to .98. Thus, placing items for the state assessments on the NAEP scale along with NAEP items is empirically justified. Results of the item mapping analyses will be presented in the final report which is under review for release in summer 2021.

Application to NAEP: This study will show how state level assessments map onto the NAEP scale. This will enhance our interpretation of NAEP results in comparison to state assessment results.

GLOSSARY

Depending on how the link is established (common items, common test takers, or randomly equivalent groups), how closely comparable the contents of the two tests are, and other considerations (e.g., the reliabilities of the compared tests or the correlation between them), one can use one of four linking procedures: equating, calibration, projection and moderation[3].

In ***equating***, both tests, *X* and *Y*, have been designed and developed to be equally reliable and each measures the same content. Equating is most often used when the goal is to relate two alternate forms of the same test, such as alternate forms of the ACT or the SAT. In equating the distributions of test *X* and *Y* are aligned or matched up directly. The matching can be done with equipercentile equating or linear equating, and the distributions can be either observed score distributions or estimates of unobserved true score distributions. Sometimes IRT scaling is applied and the resulting relationship is invariant across different populations.

In ***calibration*** (e.g., with item-response theory), two tests are assumed to measure the same content, but they are not equally reliable. For example, one test *X* might be a long test whereas the other test *Y* is short. The two versions of the test are not equated, but they are indirectly comparable because they have been calibrated to a common scale θ. This type of linking is done across years in NAEP, TIMSS, PISA, PIRLS, most state criterion-referenced tests, as well as most nationally standardized norm-referenced tests. Calibration procedures provide unbiased estimates for individual students and means (average scores), but additional statistical machinery is needed to accurately estimate group characteristics such as the variance or the percent at and above achievement levels. In the 2011 NAEP/TIMSS linking study, calibration was accomplished by scaling in the same analysis the NAEP and TIMSS items that were administered within braided (one block NAEP paired with one block TIMSS) test booklets.

In ***projection***, a regression equation uses the correlation between the two tests to predict the scores on one test *Y* from those of another test *X*. There is no assumption that the two tests measure the same content or that they are equally reliable. However, there is an assumption that the tests are highly correlated. With projection, there is no longer a symmetric relationship between one test and the other. The conversion table for predicting the first test from the second is different from the table predicting the second test from the first. A statistical link was established between the NAEP and ECLS-K grade 8 reading scales using the marginal maximum likelihood (MML) composite regression procedure with the AM software (Cohen, 2005).

In ***statistical moderation***, the scores on the first test *X* are adjusted to have the same distributional characteristics as the scores on the second test *Y*. In this case it is assumed *X* is linked to *Y*. This is typically done by matching the means and standard deviations of *X* and *Y* or by matching their percentile ranks. The usual requirement for statistical moderation is that both *X* and *Y* have been administered to comparable populations of students (e.g., the student populations taking both tests are randomly equivalent). The State Mapping Study estimated the position of each state's standards on a common scale by comparing the percentages of students in each NAEP school who achieved each of a state's performance standards with the distribution of NAEP scores by the random sample of students in the school who took NAEP.

---

[3] Phillips, G. W. (2014). *Linking the 2011 National Assessment of Educational Progress (NAEP) in Reading to the 2011 Progress in International Reading Literacy Study (PIRLS*). Washington, DC: NAEP Validity Studies, American Institutes for Research.

Status of Implementing the Achievement Levels Work Plan

Overview

Even after being in use for nearly 30 years and undergoing several evaluations, the NAEP achievement levels are still considered to be on a trial basis. The 2016 evaluation of NAEP achievement levels[1], conducted by the National Academies of Sciences, Engineering, and Medicine, focused on the NAEP mathematics and reading achievement levels for grades 4, 8, and 12. This evaluation report, presented to the Governing Board at its November 2016 meeting, stated, "During their 24 years [the achievement levels] have acquired meaning for NAEP's various audiences and stakeholders; they serve as stable benchmarks for monitoring achievement trends, and they are widely used to inform public discourse and policy decisions. Users regard them as a regular, permanent feature of the NAEP reports" (page Sum-8). This evaluation included several recommendations, and the Board issued a formal response noting its planned actions in December 2016.

One important aspect of the Board's response to the 2016 evaluation was a commitment to update the guidance provided in the Board policy statement on NAEP achievement levels. The Committee on Standards, Design and Methodology (COSDAM) began working to update this policy guidance in March 2017, and the revised policy on Developing Student Achievement Levels for NAEP was unanimously adopted by the Board in November 2018.

During the March 2019 Board meeting, Governing Board Chair Beverly Perdue established an Achievement Levels Working Group chaired by Gregory Cizek to develop a comprehensive plan (including a list of activities for the Governing Board to pursue in conjunction with the National Center for Education Statistics) to fully respond to the evaluation. Ideas were shared and discussed with NCES Commissioner Lynn Woodworth and Associate Commissioner Peggy Carr during the spring and summer of 2019, with COSDAM and Reporting and Dissemination (R&D) Committee members during the August 2019 Board meeting, and with the full Board during the November 2019 Board meeting; the Achievement Levels Work Plan was adopted unanimously by the full Board at the March 2020 Board meeting[2].

In September 2020, the Board adopted a new Strategic Vision which includes a goal to "Develop a body of evidence to improve the interpretation and communication of NAEP achievement levels to ensure that they are reasonable, valid, and informative to the public." This goal is intended to be addressed through the implementation of the Achievement Levels Work Plan.

---

[1] A free PDF of this report can be downloaded at: https://www.nap.edu/catalog/23409/evaluation-of-the-achievement-levels-for-mathematics-and-reading-on-the-national-assessment-of-educational-progress

[2] At the July 30, 2020 Board meeting, the Board unanimously voted to amend the plan slightly to remove language that resulted in the misunderstanding and misinterpretation that the Governing Board is changing its definition of *NAEP Proficient* or is in the process of revisiting the NAEP achievement levels.

Current Status of Activities (February 2021)

Less than one week after the Board adopted the Achievement Levels Work Plan on March 7, 2020, the COVID-19 pandemic resulted in massive disruption to schools and businesses across the county. Governing Board staff, NCES staff, and Board discussions pivoted to focus on the previously unanticipated priority of figuring out whether and how to conduct the 2021 NAEP assessments during the pandemic.

Some of the activities specified in the Achievement Levels Work Plan proceeded on schedule (e.g., awarding a contract for studies to review and revise the achievement level descriptions) while other activities faced some delays from the timelines that were initially anticipated. The Executive Director update in the advance materials of the May 2020 Board meeting included the following update on the implementation of the Achievement Levels Work Plan:

Achievement Level Setting (ALS) Work Plan Implementation: *The ALS work plan adopted by the Board in March 2020 is largely unaffected by COVID, with the exception of a few activities that will be challenging to begin virtually. In consultation with COSDAM Chair Andrew Ho and ALS Work Group Chair Gregory Cizek, staff determined to postpone temporarily:*

- *Convening an advisory group of technical and communication experts to provide feedback on development of materials for communicating achievement levels;*
- *Collecting information about current uses of NAEP achievement levels and evaluating appropriateness of interpretations and uses not directly intended; and*
- *Working with NCES to determine details and funding for the alignment studies of math and reading frameworks and item pools that NCES had agreed to lead.*

The postponement of these activities should not significantly impact the Board's ability to achieve the overarching goals of the work plan. We will continue to evaluate when these 3 activities can resume and/or whether alternatives for launching the work virtually need to be considered.

The tables below summarize the current status of the activities approved by the Board (by groups of recommendations). A description of how the planned activities are intended to address the recommendations from the evaluation can be found in the Achievement Levels Work Plan. In order to implement some of the planned activities (particularly those in response to Recommendations #5 and #6), a new contract vehicle will be required to support this work.

The purpose of the March 2021 COSDAM discussion is for members to ask questions and provide input on the prioritization of remaining activities.

Responding to Recommendations #1, 2, and 3 (Valid)

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.

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

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.

Activity in Approved Work Plan	**Responsibility**	**Current Status (February 2021)**
COSDAM approval of Achievement Levels Procedures Manual (described in policy statement)	COSDAM	Following discussion and agreement at the March 2020 COSDAM meeting, the Achievement Levels Procedures Manual was posted on the Governing Board website: https://www.nagb.gov/content/dam/nagb/en/documents/naep/Achievement-Levels-Procedures-Manual.pdf This is intended to be a living document and will be updated as issues and questions arise related to the implementation of the policy.
Conduct studies to examine and/or document alignment between NAEP Math and Reading Frameworks and item pools for grades 4, 8, 12	NCES	NCES has not started this work; conducting these alignment studies is contingent upon being able to convene in-person panel meetings of subject-matter experts.

Activity in Approved Work Plan	Responsibility	Current Status (February 2021)
Conduct studies to review and revise Math and Reading ALDs at grades 4, 8, and 12 Conduct studies to review and revise U.S. History, Civics, and Science ALDs at grade 8	NAGB	The Governing Board awarded a contract to Pearson in September 2020; work is currently underway with regular updates to COSDAM during quarterly meetings.
Full Board action on revised Reporting ALDs	NAGB	The studies to review and revise Math and Reading ALDs (referenced above) will produce reporting ALDs that the Board will need to approve no later than August 2022 for inclusion in the 2022 NAEP Math and Reading release. Timelines for Board approval of reporting ALDs in other subject areas will be determined by updates to the NAEP Assessment Schedule and anticipated timelines for the release of those results.
Conduct studies to examine and/or document alignment between NAEP U.S. History, Civics, Science, and TEL Frameworks and item pools for grade 8	NCES	NCES has not started this work; conducting these alignment studies is contingent upon being able to convene in-person panel meetings of subject-matter experts.

Responding to Recommendation #4 (Informative to the Public)

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

Activity in Approved Work Plan	Responsibility	Current Status (February 2021)
Review of technical memo on various ideas (including pros/cons) for synthesizing and representing findings about how other assessments and external indicators of student performance relate to NAEP (including a summary of existing linking studies) and what the findings mean for NAEP.	NAGB	This memo has been produced; it will be discussed by COSDAM as one follow-up item to the linking studies discussion at this meeting.

Activity in Approved Work Plan	Responsibility	Current Status (February 2021)
As the Governing Board works to develop its next Strategic Vision, deliberations will take place as part of that effort to determine how to approach the goal of making NAEP more relevant by connecting NAEP results to important real-world indicators of student achievement.	NAGB	The Strategic Vision adopted by the Board in September 2020 includes the following goal: "Link NAEP resources with external data sources and disseminate what is learned from these sources so that NAEP can inform policy and practice in understandable and actionable ways." Work on this Strategic Vision goal (under discussion at this meeting) should include efforts to contextualize the NAEP achievement levels by relating them to other important indicators of student achievement.
Decision on additional studies that should be pursued to connect NAEP to other assessments and external indicators of student performance	NAGB/NCES	Priorities and decisions for additional linking studies will be made as part of the implementation of the Strategic Vision goal on linking (under discussion at this meeting).

Responding to Recommendations #5 & #6 (Reasonable, Valid, Informative)

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

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

Activity in Approved Work Plan	Responsibility	Current Status (February 2021)
Convene ongoing advisory group to discuss and provide feedback on the development of materials for communicating NAEP achievement levels	NAGB/NCES	This has not yet started and will require a contract vehicle to execute this work. Planning is underway on a procurement to undertake this activity.
Collect information about current uses of NAEP achievement levels via focus groups and evaluate appropriateness of interpretations and uses that are not directly intended	NAGB	This has not yet started and will require a contract vehicle to execute this work. Planning is underway on a procurement to undertake this activity.
Adopt statement of intended purpose and meaning of NAEP	NAGB	The Board unanimously adopted a statement on the Intended Meaning of NAEP at the March 2020 Board meeting.
Improve communications of what NAEP frameworks and achievement levels represent	NAGB/NCES	This work is ongoing by the Board and NCES; future efforts are intended to be informed by the advisory group referenced above when it is convened.
Develop and finalize interpretative guide for NAEP achievement levels; iterative drafts will be discussed by COSDAM and R&D	NAGB	A technical memo was produced to suggest potential components of an interpretative guide in general (prior to any considerations about interpreting results in the context of COVID-19). The production of an interpretative guide has not yet started and will require a contract vehicle to execute this work. Planning is underway on a procurement to undertake this activity.
Collect and document validity evidence to support intended interpretations and uses of NAEP achievement levels • Collect and summarize validity evidence to support intended interpretations and uses of NAEP scale scores	NAGB NCES	This activity is intended to be a synthesis from all other activities, in addition to existing evidence; work on this summary report has not yet started.

Responding to Recommendation #7 (Valid)

Recommendation #7: NAEP should implement a regular cycle for considering the desirability of conducting a new standard setting. Factors to consider include, but are not limited to: substantive changes in the constructs, item types, or frameworks; innovations in the modality for administering assessments; advances in standard setting methodologies; and changes in the policy environment for using NAEP results. These factors should be weighed against the downsides of interrupting the trend data and information.

No specific activities were included in the Achievement Levels Work Plan in response to Recommendation #7 because it was already addressed by inclusion in the revised policy statement on Developing Student Achievement Levels for NAEP (Principle 4) and the Achievement Levels Procedures Manual.