

**National Assessment Governing Board**

**Meeting of May 18–19, 2023**

**DoubleTree by Hilton Hotel Los Angeles Downtown  
120 South Los Angeles Street  
Los Angeles, CA 90012 & Virtual**

**OFFICIAL SUMMARY OF GOVERNING BOARD MEETING**

**Complete Transcript Available**

**National Assessment Governing Board Members**

Beverly Perdue, *Chair*  
Alice Peisch, *Vice Chair*  
Alberto Carvalho  
Tyler Cramer  
Christine Cunningham  
Frank Edelblut  
Viola García  
Eric “Rick” Hanushek  
Patrick Kelly  
Anna King  
Suzanne Lane  
Scott Marion  
Reginald McGregor  
Jon Pickinpaugh  
Michael Pope  
Julia Rafal-Baer  
Ron Reynolds  
Nardi Routten  
Dilhani Uswatte  
Martin West  
Mark White  
Carey Wright  
Mark Schneider, *Ex-Officio*

**National Assessment Governing Board Members Absent**

Haley Barbour  
Russ Whitehurst

**National Assessment Governing Board Staff**

Lesley Muldoon, *Executive Director*  
Elizabeth Schneider, *Deputy Executive Director*  
Rebecca Dvorak

Stephaan Harris  
Donetta Kennedy  
Laura LoGerfo  
Munira Mwalimu  
Tessa Regis  
Sharyn Rosenberg  
Angela Scott  
Matt Stern  
Anthony White

**National Center for Education Statistics (NCES)**

Peggy Carr, *Commissioner*  
Daniel McGrath, *Acting Associate Commissioner* (delegated authority)  
Gina Broxterman  
Brian Cramer  
Enis Dogan  
Veda Edwards  
Patricia Etienne  
Eunice Greer  
Linda Hamilton  
Dana Kelly  
Shawn Kline  
Nadia McLaughlin  
Gabrielle Merken  
Emmanuel Sikali  
Michael Slattery  
Holly Spurlock  
Bill Tirre  
Ebony Walton  
Bill Ward  
John Whitmer  
Grady Wilburn  
Angela Woodard

**American Institutes for Research (AIR)**

George Bohrnstedt  
Markus Broer  
Christina Davis  
Kim Gattis  
Cadelle Hemphill  
Young Yee Kim  
Sami Kitmitto  
Kerry Vieth  
Zheng Xiaying

**Council of Chief State School Officers (CCSSO)**

Fen Chou

**Council of the Great City Schools (CGCS)**

Brian Garcia

Chester Holland

**CRP, Inc.**

Shamai Carter

Monica Duda

David Hoff

Anthony Velez

Edward Wofford

**Educational Testing Service (ETS)**

Jeff Ackley

Terran Brown

Mercy Byrne

Jay Campbell

Peter Ciemins

Gloria Dion

Patricia Donahue

Amy Dresher

Robert Finnegan

Helena Jia

Ranu Palta-Upreti

Rupal Patel

Hilary Persky

Shannon Richards

Courtney Sibley

Lisa Ward

Nancy Waters

Andy Weiss

Karen Wixson

**Hager Sharp**

Joanne Lim

Kathleen Manzo

Erik Robelen

Debra Silimeo

**The Hatcher Group**

Jenny Beard

Sami Ghani

Sophia Handel

Nandini Singh

Mallory Werthamer

**Management Strategies**

Micajah Anderson  
Brandon Dart  
Rachel Koether  
Zach Rosensteel  
Peter Sobich

**Manhattan Strategy Group**

Tara Donahue  
Lori Meyer

**Optimal Solutions Group**

Andrea Johnson  
Peter Simmons

**Pearson**

Scott Becker  
Cindy Flockhart  
Joy Heitland  
Pat Stearns  
Llana Williams

**Westat**

Margaret Bartz  
Greg Binzer  
Lauren Byrne  
Marcie Hickman  
Lloyd Hicks  
Tom Krenze  
Kavemuii Murangi  
Rick Rogers  
Megan Schneider  
Desrene Sesay  
Leslie Wallace  
Rima Zobayan

**WestEd**

Mark Loveland  
Marianne Perie  
Taunya Nesin

**Others**

Myra Best, digiLearn  
Tamika Brinson, Florida Department of Education

Oscar Carballo, Jr., Lewis-Burke Associates, LLC  
Will Donkersgoed, Wyoming Department of Education  
Karla Egan, Ed Metric  
Andrea Faulkner, North Carolina Department of Public Instruction  
David Geary, U.S. Department of Education  
Laura Goadrich, Arkansas Department of Education  
Michelle Lerner, Lerner Communications  
Rebecca Logan, Oklahoma State Department of Education  
Raina Moulain, Alaska Department of Education & Early Development  
Sara Mooney, Los Angeles Unified School District  
James Norris, Illinois State Board of Education  
Ronald Petracca, U.S. Department of Education  
Renee Savoie, Connecticut State Department of Education  
Mark Stephenson, Kansas Department of Education  
Christy Talbot, American Education Research Association  
Julie Williams, California Department of Education

### **Welcome**

Beverly Perdue, Chair, called the meeting to order at 12:39 p.m. and welcomed attendees to the May meeting of the National Assessment Governing Board (Board or Governing Board) held in-person in Los Angeles, California, and virtually.

### **Approval of the May 2023 Agenda and the March 2023 Minutes**

Perdue requested a motion for approval of the May 2023 meeting agenda. The motion was made by Marty West, seconded by Scott Marion and passed unanimously.

Perdue requested a motion for approval of the March 2023 meeting minutes. Marion moved to approve the minutes, and Julia Rafal-Baer seconded the motion, which passed unanimously.

### **Executive Director Update**

Lesley Muldoon, Executive Director of the Governing Board, thanked Alberto Carvalho for hosting the meeting and introducing the Board members to the students, teachers, and staff of the Los Angeles Unified School District (LAUSD). She credited the spoken word performance from the previous night's outreach dinner for reminding her that hope still exists in education despite academic declines. She also thanked the Governing Board staff and contractors for their work in planning and executing a Board meeting in California.

Muldoon congratulated Munira Mwalimu, who will retire in December 2023 after more than 20 years of federal service. As an executive officer, Mwalimu holds multiple responsibilities for the Governing Board, including overseeing personnel matters, IT and cybersecurity, and the budget, among other tasks. Her position will be posted soon so that a new staff member can be hired and begin work in early January 2024. Mwalimu will be at the remaining 2023 Governing

Board meetings, but Muldoon wanted to inform the Board about the transition and asked them to congratulate Mwalimu on this milestone with a round of applause.

Muldoon then began her updates by reminding the Board about the innovation agenda begun in March 2022. She noted that in May 2022, NAGB and NCES mapped out a strategy for 2022 and a set of priorities for 2023 with the goal of improving the relevance, utility, and equity of the National Assessment of Educational Progress (NAEP).

Over the past 15 months, the Board has improved its approach to Report Card releases by engaging proactively with state and local policymakers (to help them better understand and use NAEP results) as well as with media channels (to place op-eds written by Board members and alumni). Muldoon credited the attention paid to the NAEP releases this year to the sustained campaign-like approach to communicate the results rather than to traditional one-off events.

Modernization work on NAEP is proceeding. Congress approved funding for the research and development necessary in Fiscal Year (FY) 2023 to support the modernization work. The FY2024 budget request included a similar ask. Muldoon observed that the Board has not received a budget increase in approximately 15 years, but that the FY2024 budget includes a \$1.5 million increase. For FY2025, Muldoon anticipates requesting an increase to fund scheduled updates to the NAEP assessment frameworks.

Frameworks were not updated from approximately 2006 to 2018, when the Board committed to reviewing all the frameworks. The Assessment Development Committee (ADC) continues to explore how to make more frequent incremental changes rather than substantial changes every few decades. The Board is discussing how to institute these changes in a way that may maintain trend and ensure the validity of the assessments.

The Board plans to adopt an updated Science Assessment Framework in November 2023. The ADC decided during their May meeting not to update the Writing Assessment Framework. It is currently discussing options for when to next administer the assessment using the current framework. ADC will deliberate further on whether to proceed with an update to the U.S. History/Civics Assessment Framework.

Muldoon also informed the Board that they will continue the discussion about updating the language in assessment frameworks related to equity at the August 2023 Board meeting. At the March 2023 meeting, the Board began to discuss updating this language for the Science Assessment Framework.

The Board needs to make some decisions about updating the NAEP assessment schedule. The Board will decide if the 2028 Technology and Engineering Literacy Assessment should be administered and whether to move the NAEP Writing Assessment from 2030 to an earlier date.

Muldoon shared that the Board received a clean financial audit. The audit has progressed to the second phase, which is to document and have reviewed all of the Board's internal controls, processes, and procedures. The Governing Board currently has two open procurements: (1) strategic communications, which includes the Board's communications, outreach, and

dissemination activities; and (2) events and logistics. A third procurement for data visualization/multimedia support is forthcoming.

Muldoon ended her presentation by encouraging all Board members to speak up at the Board meetings. Every voice is important to help the Board make better decisions.

Perdue thanked Muldoon and the Governing Board staff for their work. She then thanked Carvalho for hosting the meeting in Los Angeles. She also acknowledged the student performers and complimented the LAUSD principals, teachers, and children the Board met.

### **IES Director Update**

Mark Schneider, Director of the Institute of Education Sciences (IES), explained that IES celebrated its 20th anniversary last year. Over the past few years, IES has aimed to modernize its research and development. To accomplish this, IES is trying to establish the National Center for Advanced Development in Education (NCADE), similar to the Defense Advanced Research Projects Agency (DARPA), which would fund research to understand what interventions in education succeed for which populations under what conditions.

A second goal of IES is to decrease the turnaround time in understanding whether an intervention will work on a large scale. Schneider said IES needs to learn “how to fail fast, and we have to learn how to learn from failure.” Scaling up is a key way to do this, but IES typically collaborates with academics who cannot scale up or commercialize their work; thus, IES should consider establishing relationships with the private sector.

IES soon will release a Request for Application to tap the expertise of technology companies through research-practice partnerships. IES also is supporting work on digital learning platforms to accelerate experimentation, research, and replication through SEERNet. IES recently established two AI Institutes with the National Science Foundation. The first one, called *AI Institute for Exceptional Education*, will focus on how to give teachers more time to work individually with students. The second one, *Invite*, will study how to use AI to improve non-cognitive outcomes.

The XPRIZE competition at IES had three winners in the first stage. However, too few applicants tried to compete for an XPRIZE in middle school science education, prompting the competition’s cancellation. Currently, IES is running an XPRIZE for elementary school math for students with disabilities.

Through the XPRIZES, Schneider learned that people are driven by data access more than the prize money. IES put out a Request for Information (RFI) to ask for help in identifying massive datasets in education. Through the RFI, only three datasets were identified, including Federal Student Aid, Statewide Longitudinal Data Systems, and NAEP. Schneider encouraged the Board to allow researchers more reliable and faster access to NAEP data so NAEP can become the “nation’s definitive source of education data.”

With the growth of ChatGPT, there is a need for authentic learning data, which is where NAEP may be used to improve learning models. Schneider suggested that ChatGPT could generate more NAEP writing prompts from sample NAEP Writing Framework questions. The prompts could be created in seconds and be subject to evaluation for inclusion on NAEP.

Tyler Cramer asked Schneider if it were possible to capture royalties from monetizing the data that could be applied to future research development. Schneider explained that IES has an obligation by law that the data collected must be made public. Dan McGrath added that NCES cannot release secure items, which is an obstacle to releasing data. This challenge could be resolved by increasing the number of items. Schneider reiterated that AI, such as ChatGPT, can generate prompts rapidly.

Rafal-Baer commented that she was excited about the possibility of the innovations and would like to learn more about how differently abled students would be supported and how these changes could help families. She also asked to learn more about if and how these innovations produce cost savings. Carey Wright seconded Rafal-Baer's comments and encouraged IES to focus on using AI to help reduce the workload of teachers who must write Individualized Education Plans for students.

Before transitioning to Carr and McGrath for the NCES Commissioner Update, Perdue congratulated Carr for being invited to give the commencement address for the School of Education at Howard University.

### **NCES Commissioner Update**

Carr said that NCES has launched its Strategic Plan, which is available on the website for stakeholders to review. The Strategic Plan sets four goals: (1) develop and disseminate innovative products aligned with today's data ecosystem; (2) improve and innovate NCES operations; (3) foster and leverage mutually beneficial partnerships; and (4) embed principles of diversity, equity, inclusion, and accessibility into all of NCES's work.

Carr also explained how successful the Pulse School Panel Survey has been and that Congress approved funding to continue administering the survey. The current survey is experimental, so NCES needs to work with the Office of Management and Budget (OMB) about changing the items. NCES is currently seeking past participants' feedback on how to improve the survey. The redesign will include the addition of the territories, and with additional funding, a district-level version of the Pulse School Panel Survey. The Pulse School Panel Survey will be administered again in August 2023.

NCES will also release the congressionally mandated Condition of Education report on May 24, 2023. NCES already briefed Congress and senior leaders in the U.S. Department of Education on the report's highlights, which include a focus on the impact of COVID-19.

OMB approved a mandate for every school and district in the country to submit finance data. Currently, all states submit it, but not every district or school has submitted finance data. This is



important data for NCES to collect and analyze to discover how schools, districts, and states use their funds.

Carr listed other assessments that have been or will be released shortly as well as upcoming linking studies. Releases include the Progress in International Reading Literacy Study (PIRLS), which was released on May 16, 2023. NAEP Long-Term Trend (LTT) will be released on June 21, 2023, and the Programme for International Student Assessment (PISA) will be released in December 2023. NAEP 2022 results will be linked to the Monthly School Survey that was collected in 2021. NAEP also will be linked to the COVID-19 School Data Hub (Hub). The Hub tracked the opening of schools and the mode of instruction.

Carr shared that in December 2022, a common standardized application process was created for anyone who wants to request data from any of the 13 statistical agencies, including NCES. NCES has received 40 new applications already. She hopes to find funding to support a staff member to manage the application process at NCES. NCES is also releasing a new data tool that allows regression analysis, analyses of gaps between different student subgroups (e.g., Black and White students), and analyses that examine what variables moderate relationships, such as between race/ethnicity and outcomes.

Dilhani Uswatte said she was excited about the regression analysis tool but wondered who was looking at the data and how user-friendly it would be. Carr confirmed it would be user-friendly (e.g., point and click).

Rick Hanushek asked how the tool compared to the work being conducted by Matthew Chingos at the Urban Institute. Chingos is running regressions on individual data and using coefficients to adjust state means. West said he did not think Chingos's work had the funding for updates and thought that the NCES tool was useful. McGrath said that he believed that the NCES tool is more limited and that NCES would not want to calculate new scores for states based on demographics.

### **Linking Studies and NAEP**

Perdue introduced the next session by explaining that West, Cramer, Rafal-Baer, Marion, and Hanushek have discussed NAEP linking studies as part of the Board's Linking Studies Working Group. She turned it over to Hanushek to lead the session.

Hanushek explained that linking NAEP data to other data sources can be invaluable by adding context to NAEP analyses and validating NAEP data. These linked datasets can address questions unable to be answered with NAEP data alone (e.g., do NAEP math scores relate to college graduation rates?). Currently, linking studies can be done at the school level by using school information included in the NAEP data. A researcher can link the datasets to various external data at the school aggregate levels, such as state assessment data or PISA. Linking can also occur at the student level, sometimes referred to as overlap sample studies; however, NAEP generally does not retain student-level information, so these studies require advanced planning and security considerations. Hanushek introduced Markus Broer to explain what can be learned through linking studies.

Broer, managing researcher at the American Institutes for Research (AIR), provided examples of both school- and student-level NAEP linking studies. He noted NAEP could be linked with NCES's Education Demographic and Geographic Estimates (EDGE) school poverty estimates, which are based on the geographic location of public schools, to inform a more precise measure of socioeconomic status (SES) for NAEP. Broer also discussed how, pre-pandemic, researchers linked Trial Urban District Assessment (TUDA) data to county-level health and socioeconomic data and found that TUDAs with the steepest score declines had the lowest SES and most adverse health outcomes, whereas TUDAs with the highest SES and better health outcomes experienced less decline.

Broer next shared findings from an analysis with an overlap sample between NAEP and the High School Longitudinal Study of 2009 (HSL:09). The HSL:09 surveyed and assessed ninth-graders starting in 2009, assessed them again in 2011 (when the cohort was in grade 11), and provided a sample of HSL:09 participants' NAEP performance in 2012 when they reached grade 12. The survey data comprise information on high school graduation, college enrollment, postsecondary attainment, financial aid decisions, etc. Broer presented data showing differences in college enrollment for students at each of the NAEP achievement levels. Students performing at *NAEP Advanced* were the most likely to attend a two- or four-year college, and they were most likely to major in a STEM subject compared to peers scoring at other achievement levels. Hanushek reiterated that this study would not have been possible without linking NAEP data to other data sources.

McGrath discussed the feasibility and challenges of student-level linking studies. To link student-level data, NCES needs OMB clearance to collect and maintain student personally identifiable information (PII). Because PII is typically not retained by NCES, the research team must plan early to ensure the appropriate data are collected correctly to allow linking NAEP to the other data source. The key lesson is that linking takes planning, and there are many data collection challenges that need to be anticipated and resolved, including subject recruitment and cost.

Hanushek concluded the session by summarizing recommendations of the Linking Studies Working Group: (1) the Board should prioritize linking NAEP student-level data to other NCES surveys, (2) the Board should consider ways to grant access to linked datasets, and (3) the Board should disseminate lessons from linked NAEP data (e.g., to provide context to achievement levels) widely.

Cramer added that this type of research is important for people who need to make decisions about resource allocation. For example, colleges may use the data to determine how many remediation programs they may need to make available. Cramer also suggested linking the length of time a student has resided in a given jurisdiction to NAEP results. If the student has not lived in the district for sufficient time, then the student's outcomes may not be attributable to that student's experience in the school where NAEP was administered.

Cramer emphasized that he would like to have the Board on record as supporting NCES's Evidence Act compliance. The Evidence Act includes the requirement that government datasets are made to be more interoperable with each other. As one of the 13 statistical agencies, NCES

has encouraged more effective and efficient use of federal datasets throughout various agencies. He would like the Board to endorse NCES's efforts.

Carr added that NCES was recently approached by the U.S. Department of Housing and Urban Development to link their data to NAEP data. She emphasized that there will be costs associated with the linking and that caution needs to be taken with conducting the analyses to prevent inaccuracies.

Marion suggested prioritizing types of linking studies. For example, people may be interested in learning more about predicting college outcomes. Schneider concurred, clarifying that IES's research and development infrastructure must become more responsive to policymakers, must inform relevant issues like workforce development, and must finish more quickly to make results more useful. Michael Pope argued that findings from linking studies can inspire students and help bolster the STEM workforce. He wondered how linking data could be used to ensure the updated Science Assessment Framework meets the needs of today's students.

Patrick Kelly expressed appreciation for the discussion and observed that if no one ever uses the linked data, what is actually accomplished? This leads to another question: What is most useful and valuable to analyze through linking studies? Kelly suggested exploring the critical lack of citizenship skills across the country and the importance of linking the U.S. History/Civics Assessment results to noneconomic factors like citizenship skills.

Ron Reynolds asked whether NAEP or the Scholastic Aptitude Test (SAT) was a better predictor of college readiness. Broer answered that both are good predictors of college readiness, especially when combined with high school GPA.

Viola García suggested that the Board consider less costly ways of disseminating data and information, especially readily accessible data. Elaborating on García's comment, Wright asked the Board to think about ways to make NAEP data more valuable to different sectors at both the federal and state levels. People outside of education, such as those in the business community, law enforcement, or health services, may want to learn from NAEP.

Perdue closed the session by thanking Hanushek and the committee for their work, Rebecca Dvorak for her leadership, and Broer for his invaluable contributions. She challenged the Board members to ask themselves what the role of the Nation's Report Card is in the dissemination of data for a broader purpose than reporting scores. What are the risks and rewards?

The session recessed at 3:28 p.m. and reconvened at 3:45 p.m.

### **NAEP Budget and Assessment Schedule (Closed)**

Under the provisions of exemption 9(B) of §552b(c) of Title 5 U.S.C., the Governing Board met in closed session on Thursday, May 18, 2023, from 3:45 p.m. to 5:45 p.m. to receive a briefing from Carr, Commissioner, NCES, and Muldoon, Executive Director.

Muldoon provided opening remarks about the importance of the Board’s role in setting policy priorities related to the assessment schedule and the budget implications from doing so. Carr and McGrath then led a briefing on the NAEP budget, which included a review of the operational costs for the program. The presentation included information on the major cost drivers for NAEP and how the adopted assessment schedule can impact costs for the program.

### **NAEP Budget and Assessment Schedule (Closed)**

Under the provisions of exemption 9(B) of §552b(c) of Title 5 U.S.C., the Governing Board met in closed session on Friday, May 19, 2023, from 9:00 a.m. to 9:30 a.m. to receive a briefing from Muldoon, Executive Director.

Muldoon provided an overview of the activities of the Ad Hoc Committee on the Assessment Schedule, which was established in November 2022 to work on proposing a solution to move NAEP assessments back to an odd-year cycle.

### **ACTION: Resolution on NAEP Assessment Schedule**

Perdue explained that this resolution asks Congress to postpone the 2026 Reading and Mathematics Assessments to 2027. The Governing Board heard concerns from across the country about the potential impact of releasing scores immediately before a federal election. The NAEP Reading and Mathematics Assessments will be given in 2024 and then again in 2027. Perdue asked for a motion on the resolution. West moved to approve the resolution and Suzanne Lane seconded. Perdue asked Pope to read the resolution into the record. After reading the resolution, Perdue called for a vote. The resolution passed unanimously.

The session recessed at 9:40 a.m. and reconvened at 9:45 a.m.

### **Discussion of Feedback on Proposed 2028 NAEP Science Assessment Framework**

Kelly began the discussion on the proposed 2028 NAEP Science Assessment Framework by introducing the four panel leadership team members, Aneesha Badrinarayan, Jenny Christian, Nancy Hopkins-Evans, and Joseph Krajcik. WestEd staff and Technical Advisory Committee member Karla Egan were also present to answer questions.

Kelly provided a brief background on the NAEP Science Framework. The framework was last updated in 2005 for the 2009 administration. A panel of science education experts has reviewed the framework to make recommendations for the 2028 NAEP Science Assessment. Diverging from previous framework changes, the Board sought initial public comment in fall 2021 about whether and how the current framework should be changed. The Board also commissioned papers from individuals in national science organizations. The strategic communications contractor for this work, which was also a new addition to the process, conducted in-depth interviews with stakeholders. This feedback allowed the panel to create an initial draft of the updated framework, which then was submitted for public feedback from mid-March to mid-April. Sharyn Rosenberg and panel members made presentations about the framework update process nationally (both at in- person and online events), in which approximately 300 people

participated; 29 individuals and organizations submitted feedback on the framework draft. NCES also provided feedback in a detailed memo. The panel plans to present a revised framework for discussion at the August quarterly Board meeting with the Board taking action at the November quarterly Board meeting.

Hopkins-Evans reminded the Board members that the National Research Council (NRC) Framework for K–12 Science Education grounded the panel’s work in research. The NRC Framework includes science and engineering practices, disciplinary core ideas, and crosscutting concepts. Together, these three dimensions inform K–12 science education and assessment.

Christian shared the panel’s initial recommendations, which included updating the construct of science achievement, updating the NAEP science disciplinary concepts and practices, integrating the crosscutting concepts, and expanding the science construct to include aspects of technology and engineering. Other recommendations included describing how the three dimensions of science would be assessed on the NAEP Science Assessment and how to report student performance in light of science-specific contextual variables. The revised framework is focused on students engaging in science instead of regurgitating isolated facts.

Krajcik spoke about how, in response to feedback from NCES about the draft framework including too much content to assess feasibly, the panel has begun to reduce the number of disciplinary concept statements by prioritizing the most important concepts. To determine the priorities, panel members asked whether the concept is central to the discipline. Is it useful in understanding the world to help us make decisions about everyday life, and is it something that is useful beyond school? Can it be measured in a meaningful way? The panel also considered grade-level progressions and concepts important across all grade levels.

Badrinarayan explained how the panel was addressing NCES’s feedback on assessment design, including the need for the framework to include discrete (standalone) items in addition to item sets and to support the development of items at the lower end of the performance scale. That support comprises drafting a complexity framework that moves from scripting to guiding to doing science. Data from state assessments based on the NRC Framework indicate that it is possible for students at all levels to respond to multidimensional science items; released items and data from the Massachusetts science assessment were shared to illustrate this point.

Badrinarayan noted that NCES had indicated that the framework draft relied heavily on item sets, but that considerable numbers of discrete items are needed to cover the framework content. She noted that the panel is working on providing additional guidance and examples to demonstrate how discrete items can be developed to reflect a phenomenon and measure at least two dimensions while maintaining a reasonable language load. Badrinarayan shared a multipart item (from the 2009 NAEP released items) about a simple phenomenon, which reflects both a practice and disciplinary concept aligned to the proposed framework.

Badrinarayan explained that the panel is also developing draft achievement level descriptions (ALDs) to describe what students should know and be able to do at the *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* achievement levels. The Technical Advisory Committee is sharing learning progressions to ground the ALDs across grade bands. Additionally, the panel is

recommending science-specific contextual variables for student, teacher, and administrator questionnaires that were omitted from the initial draft framework for public comment.

Kelly thanked the panel leadership team for their work thus far and noted that revising the framework involves a careful balance between content issues, psychometrics, and costs. He introduced Lane, the chair of the Governing Board's Committee on Standards, Design and Methodology, who elaborated on the psychometric issues. Lane described the balance between how to be cost-effective and how to ensure that NAEP sustains its stellar reputation. The Board must find the point at which feasibility and cost are optimized immediately before sacrificing the validity and quality of the assessment. This means creating a framework that can support an assessment with the fewest items and students without compromising the intended construct, which should ensure the validity of score interpretations.

Kelly also addressed feedback received from NCES that stated maintaining trend under this framework is probably not possible from a construct and technical perspective. In the initial Board charge to the panel, trend was identified as a post hoc consideration; relevance, utility, and validity were to be prioritized, but trend was considered an important goal if feasible. Kelly noted that at this juncture, it is important for the Board to affirm that there is sufficient justification for the proposed changes given the likelihood that trend cannot be maintained. Kelly reminded Board members that the NAEP Validity Studies Panel raised serious concerns with the existing framework and assessment given the recent changes in science education and assessment. In addition, there are limited state and TUDA data for science results and there is precedent for breaking trend with the last framework change in science given how rapidly the field of science changes. The public comments received were overwhelmingly in favor of prioritizing the recommended changes over the need to maintain trend. Finally, the new framework would be implemented during the 2028 administration of the assessment. The 2024 administration will collect data for one cycle of post-pandemic data using the current framework, which would allow for pre- and post-COVID comparisons with a new trend line starting in 2028.

Kelly ended by summarizing the draft policy guidance that ADC proposes to provide to the Development Panel to inform the framework revision. First, the panel should significantly reduce the volume of content to decrease the number of items and students needed without jeopardizing the measurement of the intended construct. Second, the panel should provide detailed guidance, examples, and evidence that the framework can support items at the lower end of the scale. Finally, the panel does not need to assume that the majority of existing items can be carried forward to the new assessment, even if this results in breaking trend. Kelly invited comments or questions from the Board.

Rafal-Baer expressed support and excitement for the work and asked for more information about how the lower end of performance will be measured. Badrinarayan explained that the panel recommends triangulating examples of items on existing multidimensional assessments that address the lower end of the spectrum. These items can help inform general guidance in addition to fleshing out the complexity framework. Uswatte requested that specific examples be shared in the revised draft of the framework. Hanushek added that in his view, NAEP has not adequately addressed students performing at the lower end because NAEP assessments typically have not included foundational questions. He also wondered whether grade 4 questions could be included

on the grade 8 assessment and grade 8 questions on the grade 12 assessment. Badrinarayan explained that the complexity framework includes foundational ideas. Krajcik suggested asking students to draw a model at different grade levels to demonstrate how the model becomes more sophisticated and complex as the child grows older. Lane emphasized that it is important to maintain the framework for each grade level; she would not recommend including grade 4 items on a grade 8 assessment.

Marion encouraged the panel to continue to reduce the number of disciplinary concepts beyond the initial work that was described given that this is an assessment framework rather than a set of content standards that is supposed to guide curriculum. He said that NAEP could serve as a model for states to follow on state science assessments, which often include too much content. He also urged the panel to think about ways to report results beyond life sciences, physical sciences, and earth and space sciences to reflect the changes in the new framework clearly.

Reynolds thanked the panel leadership team and ADC for their effort and asked how feasible it is to bring the development of a valid and reliable assessment into balance with cost constraints. Christian responded that it is doable, and the panel will tackle this question of balance during their meeting in early June.

Reginald McGregor added that NAEP should be part of the larger national conversation concerning new science and technology-sector jobs. McGregor spoke to a conference of industry leaders recently who look at assessments to determine what students can do. They want to recruit and hire people who can do the work and learn. The value of the NAEP Science Assessment Framework is in ensuring that the United States is moving forward to produce the workforce that protects our national security, helps grow the gross domestic product (GDP), and produces global leaders.

Marion commented that based on NCEs feedback, discrete items may help to address students performing at lower levels. He suggested continuing to review current discrete items to determine if any of them could be included in the revised framework. This would not preserve trend, but it would provide insight into changes over time. Lane agreed and added that ideally the assessment would include scenario-based tasks and well-developed discrete items to offset costs.

As a science teacher, Pope argued that the new framework should be considered an investment, not just a cost. It is an investment in the STEM crisis. If underrepresented students can attain a science identity, then those students will enjoy opportunities to succeed in science and technology regardless of their eventual scores.

Kelly returned to the slide of the proposed Board policy guidance to the panel and asked whether any Board members had any concerns or suggestions with the proposal. No comments or questions were raised, and Kelly concluded the session by noting that there was consensus on the policy guidance.

## **Member Discussion**

Perdue started the member discussion by suggesting that the Board invite AI experts to present at a future meeting to address such questions as: How does AI interject itself into our daily lives? What is the future going to look like regarding how we consume information?

She asked Board members to reflect on what their priorities are for May and June 2023. She also asked what they would like to see on future Board meeting agendas, what they would like to see changed, and/or ideas on how to form different collaborations. Specifically, Perdue argued for stronger collaboration between the Governing Board, NCES, and IES.

Kelly recommended the Board continue to do outreach to a broader community of stakeholders so they interpret NAEP data correctly. He said the Reporting and Dissemination Committee has done an excellent job with the releases, and it needs to continue as NAEP gains more public awareness. Kelly also reiterated the importance of focusing on the Civics Assessment outcomes, adding that NAEP results should demonstrate that students who score at *NAEP Proficient* or *NAEP Advanced* have better social outcomes.

Wright shared Perdue's interest in learning more about AI, including its possibilities, challenges, and costs. Lane echoed Wright's comments and suggested that Board members review extant work with automated item generation using AI, particularly in the medical field.

West mentioned that after the LTT release in June 2023, there will be a gap of approximately 18 months when no NAEP data will be released. He encouraged the Board to think about the types of events that they could do to maintain the visibility of NAEP. Hanushek agreed and remarked that there are many studies, including linking studies, which exist but are not well-known. There may be opportunities for the Governing Board to review and share those studies.

Board members appreciated the time spent with the Los Angeles students. For example, Uswatte valued visiting with the students and asked the Board to consider implementing a student advisory board. The Board could ask students about ChatGPT to learn more about its possibilities and any concerns students may have. Nardi Routten commented that after visiting the schools yesterday, she believes in the future. She also reiterated several earlier discussion points, such as devoting more time to learning about AI and ChatGPT. She also reminded the Board members not to underestimate any student.

Reynolds added that an XPRIZE could be established for students to develop innovative ideas. Schneider responded to these ideas and said they had thought about datathons for students but were unable to implement them due to several factors. Schneider also reminded the Board that even with AI-generated items, humans still need to evaluate the questions and prompts and be involved in the development process.

McGregor thanked the Nominations Committee and encouraged Board members to think about their networks and who may want to serve on the Board. He also thanked Carvalho for his time serving on the Board. He recommended that Board members reach out to their senators and U.S. representatives, particularly those who serve on education and workforce committees because the



Board's voices matter to them. He added that Board meeting agendas should include enough time for all members to share their input and have their questions answered. Presentations are important, but the Board members need time to do the work they are charged to do. Rafal-Baer agreed and specifically mentioned that follow-up is needed from the closed session on the NAEP budget and assessment schedule. Marion agreed and recommended more small group discussions.

Perdue reminded the Board that they have the power to appoint working committees. If there is a topic that needs deeper discussion or focus, working committee suggestions should be made to Muldoon.

Carvalho reflected on his Board experience. Over the past eight years, he has pushed to use NAEP data to improve practice, but the Board continues to fail to make meaningful connections with practitioners at all levels, especially at the state-level. Innovation is not just about the use of technology but also about how to maintain active engagements and conversations with the broader public. He reminded the members that ChatGPT is a revolution, but it is likely not going to be a panacea to solve everything. Across Los Angeles, Carvalho is leading an initiative to create Individual Acceleration Plans (IAPs) for all students. IAPs will assess students in real-time and provide important insights into their learning needs without using a traditional assessment.

Pope mentioned a keynote he gave at Disney World called *Education Frozen in Time* that discussed how education was designed for more affluent males when he began his career. This can still be the case, so he encouraged the Board members to consider changes in meeting the needs of all students. He also said that the teacher shortage issue is something that the Board needs to consider as the issue becomes more pressing.

The session recessed at 11:42 a.m. and reconvened at 1:17 p.m.

**WORKING LUNCH: Briefing on Results from 2023 NAEP Long-Term Trend – Age 13 (Closed)**

Results were embargoed until June 21, 2023, so the session was closed to ensure the integrity of the NAEP program and maintain the embargo on the unreleased results.

**ACTION: Release Plan on NAEP Long-Term Trend – Age 13**

Because he had not been at the last Reporting and Dissemination Committee meeting, Carvalho asked West to present the LTT release plan. West explained that a Governing Board member will provide an introduction at the event on June 21, 2023, and NCES will be responsible for presenting the LTT results. The event will be held at a school in Frederick, Maryland. A panel of three teachers (i.e., English language arts, math, and social studies) and a student panel will share strategies to accelerate learning. This event is the capstone of a series of post-pandemic event releases. This is an opportunity to consider the results of all four assessments administered post-pandemic to provide an opportunity to look forward. Although no new releases are scheduled for

the next 18 months, the Board can continue discussions and sustain the momentum of attention developed over the past year.

Cramer moved to approve the release plan on NAEP LTT – Age 13. Wright seconded the motion, which passed unanimously.

### **Measuring Socioeconomic Status (SES)**

Broer and William Ward from NCES discussed measuring SES. Broer presented his research conducted with co-authors Yifan Bai and Rebecca Shipan on a new composite SES index and how it functions. Economic disadvantage on NAEP is currently defined as a binary eligible/not eligible for the National School Lunch Program (NSLP), which limits its validity. Broer's research argues that NAEP can transcend the binary nature of this variable and add other economics-related variables to develop an SES index, with greater variability, comprising four components: (1) number of books at home, (2) student eligibility for NSLP, (3) percentage of students eligible for NSLP at the school the student is attending, and (4) the highest level of education of either parent. Broer stated that NAEP has all the ingredients needed to construct this SES index. The proposed SES index performs well against established criteria and compares well to similar SES indices and similar assessments administered in the United States.

This index will become a categorical variable or variables available on the Nation's Report Card as part of the NAEP data tool, similar to other indices and variables. To succeed, four issues with the index must be addressed: (1) replicating Broer's results, (2) dealing with missing data, (3) developing interpretive guidance, and (4) evaluating trends with the index. The goal is to release the SES index as a new variable in December 2024.

Ward also talked about exploring the use of geospatial data to inform future SES measures. NCES's EDGE is an indicator comprising demographic and geographic estimates. Among other things, EDGE uses spatial data to create geographic location indicators, school point locations, and other information to support spatial analysis. EDGE provides income-to-poverty Ratio (IPR) values for schools. NCES could work with school districts to create similar estimates, using geocoding, for NAEP students without collecting and retaining the students' addresses. NCES may try this approach with a few districts to determine the feasibility of collecting this information.

West explained to the Board that this work has been the result of a multi-year effort that the Reporting and Dissemination Committee has been following. He asked Ward how NCES will address the six states and private schools that do not have the SES index available because they do not participate in the questionnaires. Ward responded that states do not have to participate in the questionnaires, but he hopes that the improved data will motivate them to do so. West raised a concern that the NSLP percent is a measure of school SES, and our ultimate goal is to understand this at the student-level. He wondered if an index without the school-level information is being considered. Broer understood the concern but noted the school-level variable works well because there is some similarity between the family and the neighborhood they live in. He suggested that a three-component index be created that does not include the

NSLP so both versions would be available to researchers. West urged investigating a measure of student income, something NAEP does not currently capture.

Frank Edelblut raised some concerns. First, his state is legislatively prohibited from participating in contextual variable activity. Second, he noted that Supplemental Nutrition Assistance Program eligibility recently increased in his state, which demonstrates these measures can be dynamic. Broer acknowledged that the index components are subject to change over time, but the measures also constantly undergo verification and validation processes.

Marion inquired about the use of coding the NSLP variable as 0 and 3 versus 0 and 1. Broer noted it would probably make sense to code as 0 and 1 if it helps with the interpretability. Marion then argued that aggregate data can be effective and that NAEP inferences are generally made at the larger school level and/or at the state level. He then asked Ward if the IPR would be added to the SES index or be left as a separate measure. Ward responded that their goal is to marry it to the index to see how it works. Broer noted that the IPR index worked pretty well within the index, though the percent SES variable worked slightly better. Cramer confirmed with Ward that the IPR data does not go back to 2003. Ward believed it would be available starting in 2018 or 2019, but they are updated annually.

Christine Cunningham asked about the books in the home variable and how to factor in digital books. Even if students are introduced to books, they may not be physical copies. Carvalho and Wright also echoed that question. Uswatte pointed out that young children may not be able to differentiate their own books from their parents' books. She asked how state policies that purchase books for students are factored into the analysis.

Wright also suggested that any studies done with a TUDA should also have a parallel study run in a non-TUDA school district to note any differences. She recommended learning more about each state's redistricting policies to understand if there may be concerns with estimating student SES based on a neighborhood indicator. Uswatte asked how state redistricting would impact the data. Ward said that all these factors must be considered. McGrath and Ward explained that the American Community Survey provides non-economic data. Ward also said that EDGE has collaborated with 15 states to collect longitudinal data with the geocode, which makes them confident they can do it with TUDAs as well. Carvalho added that there are some concerns about using geographic data, especially in a district like LAUSD, where local schools may not serve the students in their immediate communities.

McGregor spoke to concerns about discussing poverty. People often have pride in their communities and schools, so some of this information needs to be communicated carefully. He asked why this information is important for the Governing Board. West responded that this type of research can determine whether educational outcomes have become more or less equitable over time and how that has been influenced by policy choices over the past several decades. NAEP is one national data source to address that question. García affirmed that these questions are critical to examine equity.

### **Concluding Remarks and Student Performance**

Carvalho concluded the meeting. He hoped that the Board members left with lasting memories of their experience in witnessing the students of Los Angeles along with the teachers and administrators. He then introduced Sara Mooney, the LAUSD liaison to the Governing Board, who arranged the events. She thanked the Board members for coming to Los Angeles. She introduced two graduating seniors from Torres East LA Performing Arts Magnet program and their teacher. She explained that the students had experienced readings of Langston Hughes poems set to jazz at Huntington Gardens, and for today’s presentation, the students would recite their own poems as a response to the Hughes poems they had heard.

At the end of the student performance, Carvalho thanked the Board members for allowing him to share the Los Angeles experience with them—the good, the beautiful, and the real.

The meeting adjourned at 2:43 p.m.

*Beverly E. Perdue*

07/17/2023

\_\_\_\_\_  
Beverly Perdue, Chair

\_\_\_\_\_  
Date

**National Assessment Governing Board**  
**Executive Committee Meeting**  
**Report of May 4, 2023**

**CLOSED SESSION**

**Executive Committee Members:** Beverly Perdue (Chair), Alice Peisch (Vice Chair), Christine Cunningham, Patrick Kelly, Martin West.

**Executive Committee Members Absent:** Haley Barbour, Alberto Carvalho, Suzanne Lane, Reginald McGregor, Carey Wright.

**National Assessment Governing Board Members:** Tyler Cramer.

**National Assessment Governing Board Staff:** Lesley Muldoon (Executive Director), Elizabeth Schneider (Deputy Executive Director), Rebecca Dvorak, Stephaan Harris, Laura LoGerfo, Munira Mwalimu, Sharyn Rosenberg, Angela Scott, Matthew Stern, and Anthony White.

**National Center for Education Statistics (NCES):** Peggy Carr (Commissioner), Daniel McGrath (Delegated Authority of Associate Commissioner), Tammie Adams, Jing Chen, Brian Cramer, James Deaton, Enis Dogan, Veda Edwards, Patricia Etienne, Eunice Greer, Dana Kelly, Shawn Kline, Gabrielle Merken, Emmanuel Sikali, Holly Spurlock, William Tirre, Bill Ward, Ebony Walton.

**U.S. Department of Education Staff:** James Forester (Office of Legislation and Congressional Affairs), Brittany Beth (Budget Office).

The Executive Committee met in closed session on May 4, 2023, from 3:00 p.m. to 4:30 p.m. to receive a briefing on the program's budget and assessment schedule, the congressional activity to reauthorize the NAEP law, and ongoing work.

The Honorable Beverly E. Perdue, Chair, presided over the meeting. Perdue welcomed members and spoke about the Board's policy priorities related to the program's budget. She then invited Lesley Muldoon, Executive Director, and Matthew Stern, to provide an update on recent congressional activity related to the reauthorization efforts of the Education Sciences Reform Act.

Peggy Carr, Commissioner of the National Center for Education and Dan McGrath, Acting Associate Commissioner of NCES, then provided an update on the NAEP program's budget, projected funding flows, and considerations for the Assessment Schedule.

Stern then provided an update on the Ad Hoc Committee on the Assessment Schedule's deliberations regarding the NAEP schedule in 2026 and the effort to postpone the assessments until 2027.

These discussions were conducted in closed session because the disclosure of budget and spending plans would significantly impede implementation of agency actions as it relates to contract actions. Therefore, this discussion is protected by exemption 9(B) of section 552b(C) of Title 5 U.S.C.

At 4:30 p.m., Chair Perdue adjourned the meeting.

I certify the accuracy of these minutes.

*Beverly E. Perdue*

---

Beverly Perdue, Chair

08/03/23

Date

# National Assessment Governing Board

## Assessment Development Committee

### Report of May 17, 2023

#### OPEN SESSION

**Assessment Development Committee Members:** Patrick Kelly (Chair), Christine Cunningham (Vice Chair), Frank Edelblut, Viola Garcia, Reginald McGregor, Jon Pickinpaugh, Nardi Routten, and Dil Uswatte.

**Assessment Development Committee Members Absent:** None

**National Assessment Governing Board Members:** Scott Marion.

**National Assessment Governing Board Staff:** Stephaan Harris, Lesley Muldoon (Executive Director), Elizabeth Schneider (Deputy Executive Director), and Sharyn Rosenberg.

**National Center for Education Statistics (NCES):** Enis Dogan, Eunice Greer, Dana Kelly, Nadia McLaughlin, Holly Spurlock, and Bill Ward.

**Other attendees: American Institutes for Research (AIR):** Kerry Veith. **Educational Testing Service (ETS):** Terran Brown. **WestEd:** Mark Loveland and Taunya Nesin.

The Assessment Development Committee (ADC) met in open session on Wednesday, May 17<sup>th</sup> from 10:30 a.m. – 12:30 p.m. PT.

#### **Discussion of Consultant Papers on 2030 NAEP Writing Assessment Framework**

Patrick Kelly called the meeting to order at 10:30 am and thanked everyone for traveling to Los Angeles early to attend the ADC meeting. He noted that the Committee would be spending most of its time on various topics related to NAEP frameworks, followed by item reviews.

Kelly reminded Committee members that the NAEP Assessment Schedule indicates that the Board will consider updates to three frameworks for the 2030 administration: Writing, U.S. History, and Civics. ADC had begun discussing the Writing Framework first last fall to avoid having to update three frameworks simultaneously. The Board solicited public comment between late November and late January on whether and how the existing NAEP Writing Framework should be changed, and NCES provided a memo of operational considerations. Five consultants were asked to write short papers on this topic, which were included in the ADC materials.

Kelly noted that the Committee had also discussed budget considerations related to this work, since the current Governing Board budget does not include enough funds to pursue all of the planned framework updates. Additional funds for this work had been requested in the Board's Fiscal Year 2023 budget but were not received. The Board's Fiscal Year 2024 budget will have implications for potential framework updates; this allocation is not yet known but the President's budget request for the Governing Board was lower than what the Board staff had requested.

Kelly explained that the purpose of the current discussion was to determine whether the Committee believes that updating the NAEP Writing Framework is necessary and advisable at this time. If yes, then the Committee needs to prepare a recommendation on the degree of substantive change to the framework that is anticipated. If not, then the Committee needs to discuss implications of this decision for administration of the next writing assessment (considering both technical and budget issues).

Kelly stated that all five expert consultants noted the potential impact of artificial intelligence (AI) on writing instruction and assessment, although they differed on whether they advised the Board to hold off on updating the framework or proceed anyway. Kelly explained that after speaking with ADC Vice Chair Christine Cunningham and Board staff, he is recommending that this is not the right time to undertake an update to the NAEP Writing Framework. A new framework could be outdated quickly, potentially even prior to the first administration of the assessment. In addition, there is a concern that updating the framework in the midst of active debates in the field on the role of AI for writing assessment could put the Governing Board in a precarious position of taking a stand before there is emerging consensus in the field.

ADC members discussed the proposed recommendation. They agreed that it is important for NAEP to continue to assess writing, but that it is better to hold off on updating the framework for another few years. Instead, they expressed interest in administering the existing assessment as soon as feasible, which could inform the framework update in addition to providing data on how students are writing before AI has a larger impact. Kelly noted that the Committee should discuss considerations related to re-administering the existing writing assessment during the August Board meeting.

### **Potential Updates to Assessment Framework Development Policy**

Cunningham noted that the next agenda item was an update on efforts to explore the feasibility of making smaller, more frequent changes to NAEP assessment frameworks in the future. Following a May 2022 ADC discussion on this topic, Board staff commissioned expert papers from six consultants to respond to the questions that Sharyn Rosenberg had laid out in a framing paper. The papers were discussed at the November 2022 ADC meeting, and Rosenberg then facilitated a technical panel discussion of the consultants at the end of January. Detailed minutes from that meeting were sent to ADC members prior to the March 2023 ADC meeting and were included in the advance materials for this meeting.



Rosenberg briefly summarized the goal of this effort. Currently, about every 10 years, the Board asks whether changes to a given assessment framework are needed. Although the range of potential answers to this question are: “No,” “Minor updates,” and “Major updates,” the length of time between revisions means that the default answer will be “Major updates.” In thinking about a potential new framework development process, the goal would be to ask on an *ongoing basis* whether changes to a given assessment framework are needed. In this scenario, even when substantive changes are needed, the default would likely be more limited rather than extensive changes.

Rosenberg noted that additional clarification is needed around the process and frequency for evaluating frameworks on an “ongoing basis” and an abbreviated process for recommending more limited changes to frameworks. She explained that prior discussions have primarily focused on the latter but that there may be great benefit in having a more systematic process for monitoring frameworks on a regular basis, including but not limited to the ability to make smaller changes on a more frequent basis.

Currently, the work that precedes the official launch of a framework update is done on an ad hoc basis; panels of experts are not convened until after the Board issues a formal charge and a contract is awarded. However, much of the initial work (e.g., research on how a NAEP framework compares to state standards, public comment on whether and how the current framework should be changed, consultant papers, panelist nomination process) could benefit from oversight by experts knowledgeable about a current NAEP framework and content and policy issues in a given subject. Content Advisory Groups could engage in a coherent and systematic process for monitoring changes to a field and potential implications for NAEP frameworks. These groups could oversee and synthesize the “pre-work” that precedes an official framework launch, make initial recommendations to the Board about whether and how a framework should be updated, and continue to play a role in carrying out the updates. For example, minor changes could be carried out by the Content Advisory Group but for major changes, a subset of Content Advisory Group members could serve on the Development Panel (possibly in a leadership position).

Rosenberg briefly described the changes that would be needed to the current policy to reflect this proposal: a new principle to describe the framework monitoring process; a clear distinction between “minor updates” and “major updates”; a new principle to describe how minor updates should be carried out; and elimination of references to the Steering Panel (if major changes are carried out by a Development Panel with oversight from some or all members of the Content Advisory Group).

Rosenberg then proposed that a Social Studies Content Advisory Group be convened this fall to both serve as a proof of concept and provide advice to ADC and Board staff on preparing for updates to the U.S. History and Civics frameworks. This group could provide input on what information and research to gather to inform the update and how to navigate content, policy, and other issues. Since the current Board policy is silent on how the pre-work will be carried out, this could serve as an opportunity to try out a new approach for the initial stages of the work before fully committing to changing the policy.

ADC members expressed strong support for the idea of convening a Social Studies Content Advisory Group to both pilot the concept and inform the prework for the U.S. History and Civics frameworks. They requested additional information on the specific tasks to be performed by this group, and Rosenberg suggested that she share a draft statement of work with the Committee for review in August. ADC members stated that it is premature to markup the current assessment framework development policy in tracked changes until more information is learned from trying out the Content Advisory Group process.

## **Policy Guidance on Proposed 2028 NAEP Science Assessment Framework**

Kelly introduced WestEd project co-director Mark Loveland and science content lead Taunya Nesin to provide a general project update and preview some of the Development Panel work that would be shared in the plenary session the following day.

Nesin began by reminding ADC members of the role and importance of the National Research Council (NRC) Framework for K-12 Science Education and its relevance for the 2028 NAEP Science Assessment Framework. This evidence-based framework reflects current research in science education and assessment and consists of three interrelated dimensions: disciplinary core ideas, science and engineering practices, and crosscutting concepts. Since 2005, most states have revised their science standards to reflect sophisticated expectations of understanding science ideas and practices, including: application of the three dimensions to make sense of phenomena and solve complex problems; inclusion of technology and engineering; centrality of phenomena and problems based in real-world contexts; emphasis on sense-making that integrates the three dimensions of science; and focus on learning challenging ideas across time.

Nesin then reviewed the initial recommendations from the Steering Panel: (1) update the construct of science achievement; (2) update disciplinary concept statements and practices, and add crosscutting concepts; (3) expand the construct to include aspects of technology and engineering; (4) describe how the three dimensions of science (disciplinary concepts, science and engineering practices, and crosscutting concepts) should be assessed; and (5) describe how student achievement should be reported with science-specific contextual variables.

Nesin explained that the Development Panel met virtually on May 2<sup>nd</sup> to review the feedback received during public comment and from NCEES' review. One of the most important pieces of feedback that the panel discussed was the need to significantly reduce the volume of content in the framework to address concerns about cost and feasibility. Since that meeting, panelists agreed on prioritization rules for reducing the content to ensure that this exercise was carried out in a consistent manner across the disciplinary concept domains.

Loveland described the current status of work on assessment design, including: adding details around item types; providing specific guidance/examples and empirical evidence of multidimensional items working at the lower end of the scale; providing guidance for how to select phenomena/problem situations without introducing unintended biases;

and providing specific guidance/examples for developing multidimensional discrete items.

Loveland shared a draft complexity framework that the panel is using as a starting point for discussions on how to develop items across the full range of student performance, including the lower end of the scale. The complexity framework begins with scripted integration (where an item includes a well-defined set of actions or procedures to apply practices and crosscutting concepts to disciplinary concepts), proceeds to guided integration (where guidance is embedded in the item to support students in applying practices and crosscutting concepts to disciplinary concepts) and culminates with doing science (where student how to engage with the disciplinary concepts, practices, and crosscutting concepts).

Loveland noted that the working draft of the framework that went out for public comment focused heavily on item sets, but that the panel is working on providing examples and specifications of discrete items based on successful implementation in other assessment programs. He described additional work that is currently underway, including development of draft achievement level descriptions and guidance for subject-specific contextual variables.

Kelly then explained that the Board needs to provide policy guidance to the Development Panel as they revise the framework in response to the feedback received, and the full Board will discuss the proposed policy guidance during a plenary session on Friday morning. He described the need to balance cost and feasibility with validity and quality to identify the point at which the intended construct can be adequately measured at a reasonable cost. The public comment draft framework was too heavily weighted towards validity and quality, with insufficient attention to cost and feasibility issues. The large volume of content and heavy emphasis on set-based items would result in the need for too many items and students.

On the other hand, Kelly noted that it is important not to neglect considerations of validity and quality when trying to minimize costs. For example, requiring that the majority of existing items be carried over to the new assessment even if they do not measure the intended construct or eliminating set-based items altogether would threaten the validity of the assessment inferences. Kelly proposed that a careful balance of content, measurement, and cost issues should result in a framework that is representative but not exhaustive of important content, and that includes a combination of discrete and set-based items but limited scenario-based tasks (the most expensive type of set-based items).

Kelly then pointed out that NCES had indicated it is unlikely that trend will be able to maintained from the current assessment to the new framework, given the extent of changes proposed to the construct. Kelly reminded ADC members of the bullet in the Board charge related to trend, which was described as “important if possible” but not at the expense of validity, utility, and relevance. He stated that the proposed changes to the framework which threaten trend are central to the intended construct rather than incidental changes; therefore it is not possible to make minor changes that would

greatly increase the probability of maintaining trend. He reviewed several reasons why breaking trend in this specific situation is justified, including the initial concerns raised about the current framework by the NAEP Validity Studies Panel; the long amount of time (18 years) since changes were last made in a rapidly-changing discipline; limited state and TUDA data from the current assessment; precedence for starting a new trendline with the previous framework change in science; the compelling nature of the public comments related to implementing the suggested changes versus constraining changes to maintain trend; and the opportunity for the 2024 administration of the current assessment to provide a post-COVID data point for science before the new assessment is administered in 2028.

Kelly ended by explaining that Cunningham will attend the next Development Panel meeting on June 5-6 and provide policy guidance to the panel. He proposed the following guidance for ADC discussion: (1) significantly cut down on the volume of content to reduce the number of items and students as much as possible without jeopardizing the measurement of the intended construct (the Technical Advisory Committee should work closely with the Development Panel to provide guidance for identifying this balancing point); (2) provide detailed guidance, examples, and evidence that the framework can support items at the lower end of the scale; and (3) the panel does not need to assume that the majority of existing items can be carried forward, even if this ends up leading to a break in trend.

ADC members expressed support for the proposed policy guidance and for the current direction of the framework focusing on knowledge-in-use. They requested additional information on how the complexity framework might be used with specific examples. They also noted the importance of ensuring that the reading load and linguistic complexity does not hinder performance. The Committee was supportive of the likely need to begin new trend lines based on the updated framework but reiterated the importance of reminding stakeholders that the 2024 administration of the last assessment under the current framework will provide an opportunity to compare science achievement from before and after the COVID-19 pandemic.

## **CLOSED SESSION**

**Assessment Development Committee Members:** Patrick Kelly (Chair), Christine Cunningham (Vice Chair), Frank Edelblut, Viola Garcia, Reginald McGregor, Jon Pickinpaugh, Nardi Routten, and Dil Uswatte.

**Assessment Development Committee Members Absent:** None

**National Assessment Governing Board Members:** Vice Chair Alice Peisch and Chair Bev Perdue.

**National Assessment Governing Board Staff:** Lesley Muldoon (Executive Director), Elizabeth Schneider (Deputy Executive Director), and Sharyn Rosenberg.

The Assessment Development Committee (ADC) met in closed session from 1:45 – 2:15 p.m. to discuss budget considerations for framework updates. This session was closed because it included information about federal budgets that has not been released to the public.

### **Budget Considerations for Framework Updates (CLOSED)**

Kelly noted that the NAEP Assessment Schedule indicates that framework updates will be considered for several subjects in the coming years. Prior to the recent updates of Reading, Math, and Science, it had been many years since NAEP frameworks were updated. He described cost considerations and implications due to framework development for both the Governing Board and NCES.

ADC members discussed initial assumptions for framework and item development in Science, Writing, U.S. History and Civics to inform the federal budgeting process.

### **CLOSED SESSION**

**Assessment Development Committee Members:** Patrick Kelly (Chair), Christine Cunningham (Vice Chair), Frank Edelblut, Viola Garcia, Reginald McGregor, Jon Pickinpaugh, Nardi Routten, and Dil Uswatte.

**Assessment Development Committee Members Absent:** None

**National Assessment Governing Board Staff:** Sharyn Rosenberg.

**National Center for Education Statistics (NCES):** Eunice Greer, Dana Kelly, Nadia McLaughlin, and Holly Spurlock.

**Other attendees:** **American Institutes for Research (AIR):** Kim Gattis and Kerry Veith. **Educational Testing Service (ETS):** Terran Brown, Jay Campbell, Hilary Persky, and Karen Wixson.

The Assessment Development Committee met in closed session from 2:15 – 3:30 p.m. to review proposed assessment items for 2024 pilot administrations of the mathematics and reading assessments at grades 4 and 8. This session was closed because it contained secure items that have not been released to the public.

### **Review of 2024 Pilot Mathematics and Reading Items at Grades 4 and 8 (CLOSED)**

Cunningham thanked ADC members for sending comments on the items to Rosenberg in advance of the meeting to maximize efficiency of the meeting time. She and Kelly then reviewed each of the comments received and ADC members discussed whether items should be retained, modified, or dropped for the pilot administrations. NCES staff

and contractors displayed the relevant items during discussion and answered Committee member questions.

Following the meeting, the official ADC comments were submitted to NCES.

The meeting adjourned at 3:30 p.m.

I certify the accuracy of these minutes.



\_\_\_\_\_  
Patrick Kelly, Chair

June 8, 2023

\_\_\_\_\_  
Date

**National Assessment Governing Board**  
**Committee on Standards, Design and Methodology**  
**Report of May 25, 2023**

**OPEN SESSION**

**Committee on Standards, Design and Methodology (COSDAM) Members:** Suzanne Lane (Chair), Carey Wright (Vice Chair), Eric Hanushek, Alice Peisch, Michael Pope.

**COSDAM Members Absent:** Scott Marion, Russ Whitehurst.

**National Assessment Governing Board Staff:** Sharyn Rosenberg; Elizabeth Schneider.

**National Center for Education Statistics (NCES):** Peggy Carr (Commissioner), Daniel McGrath (Delegated Authority of Associate Commissioner, Assessment Division), Jing Chen, Enis Dogan, Eunice Greer, Nadia McLaughlin, Gabrielle Merken, Ebony Walton

**Other attendees:** **American Institutes for Research (AIR):** Brittany Boyd, Markus Broer, Cadelle Hemphill, Young Kim, Sami Kitmitto, Xiaying Zheng; **Arkansas Department of Education:** Laura Goadrich; **CRP:** Shamai Carter, David Hoff, Anthony Velez; **Educational Testing Service (ETS):** Terran Brown, Jay Campbell, Amy Dresher, Kadriye Ercikan, Jia Helena; **Hagar Sharp:** Kathleen Manzo, Erik Robelen; **IES:** Angela Woodard. Management Strategies: Mike Anderson, Brandon Dart, Zach Rosensteel, Rachel Koether; **Optimal Solutions:** Peter Simmons. Pearson: Paula Rios; **P20 Strategies LLC:** Andrew Kolstad; **Westat:** Greg Binzer, Lauren Byrne, Tom Krenzke, Rick Rogers, Keith Rust, Leslie Wallace.

**Welcome and Updates**

The Committee on Standards, Design and Methodology met (in open or closed?) session on Thursday, May 25, virtually via Zoom from (insert start – end time and time including am/pm and time zone). Suzanne Lane (Chair) began by welcoming the four COSDAM members initially on the call, and by noting that Alice Peisch planned to join the meeting late. Scott Marion and Russ Whitehurst were not able to attend.

Lane next presented key outcomes from the March 2023 COSDAM meeting, two recommendations made following an NCES presentation on adaptive testing. First, COSDAM members stressed the importance of including analyses to examine whether there are differential impacts of adaptive testing by student ability level. Second, they recommended communication around why adaptive testing is comparable to linear testing for the purposes of continuing trends.

Next, Lane reminded the group that following the March 2023 joint COSDAM and Reporting and Dissemination (R&D) meeting, the group agreed a priority should be placed on increasing the number of items at the low-end of the achievement scale. It also agreed that it is not a priority to pursue a new achievement level at this point though this could be revisited in the future.

Lane provided an update on the Achievement Level Description (ALD) Study that was currently wrapping up. The contract end date for the work is June 15, 2023, and a final technical report for the U.S. History, Civics, and Science study will be available later this summer. She noted that Board policy calls for new Reporting ALDs following the administration of new frameworks or periodically, and so additional similar studies will be needed in the future.

Lane described topics under consideration for upcoming COSDAM meetings – including a deep dive presentation by NCES staff into plans towards a device agnostic NAEP administration and a presentation on how power analyses are used for identifying NAEP sample sizes.

## **Follow-up Discussion: NAEP State-Level Sample Sizes and Effect Sizes**

The next item on the agenda was a follow-up on discussions regarding the use of effect sizes, in addition to statistical significance testing, as potential ways to address recent decreases in state-level sample sizes. Significance testing is a method of determining if differences or changes in scores are meaningful, given a certain confidence level (e.g., 95%). Significance testing does not indicate the size of the difference and is sensitive to sample size. COSDAM members had previously expressed that effect sizes may help provide additional meaning to score differences, and could be useful for reporting along with significance testing for significant and non-significant findings.

Lane provided general background on statistical significance and effect sizes for those without statistics or psychometric backgrounds. She thanked Enis Dogan of NCES for providing information prior to the meeting that summarized the purposes and interpretations of significance testing and effect sizes. For additional background, Rebecca Dvorak (Assistant Director for Psychometrics) walked the group through examples of effect size calculations using recent NAEP mathematics data.

Lane next initiated a discussion, requesting input from COSDAM members on their thoughts about: (1) whether the current NAEP standard indicating effect size should only be considered for statistically significant results and (2) the use of effect sizes in addition to significance testing in general. She noted that COSDAM members had a head start on this discussion – through an email thread the days leading up to the meeting wherein they questioned the NAEP standard that recommends considering the use of effect sizes for statistically significant findings only.



Rick Hanushek expressed that it is difficult to understand the meaning of score differences from score differences alone but that reporting differences along with effect sizes can make comparisons across subjects more interpretable. He added that, in his experience, highlighting an eight-point score difference on NAEP to emphasize learning loss is meaningless to many, but that effect size has known meaning. Currently NAEP does not include standard deviations or effect sizes on the report cards. Hanushek noted that effect sizes are often included in reporting outside NAEP regardless of significance testing. Daniel McGrath (Delegated Authority of Associate Commissioner for NCES) agreed that effect sizes could aid with understanding score differences, but did not feel it was in line with the NAEP standards to include them to add meaning to non-significant findings. Dogan added that he was concerned that it could be confusing if an effect size indicates a meaningful difference, but the significance test shows non-significance. Alternatively, there could be situations when a small effect size is associated with a significant difference. Lane understood the argument; however, thought it was still worth considering further.

Michael Pope noted that as a practitioner, he would ultimately defer to statisticians and psychometricians for the appropriateness of reporting effect sizes. He expressed that the Board should consider how to report numbers in a manner that will help address educational needs, including how teachers can use the information to meet their students' needs. Carey Wright expressed a similar sentiment, noting it is important to consider what is the bottom line and why is it important. The average person should be able to understand what is reported.

In regard to concerns about public understanding, Hanushek noted that the public also has limited understanding of statistical significance and how to interpret differences between score points. Noting the score difference as a point difference only has no meaning, including to psychometricians. Wright added that it can be confusing when a three-point difference in one state is statistically significant, but the same size difference in another state is reported as not meaningful. She expressed it could be useful to find ways to add clarity to the meaning of score differences.

Hanushek inquired about whether the Board could produce a separate report from the NCES report card, and Lane thought this was worth considering. McGrath expressed concern this could lead to confusion. Dogan offered a compromise – that the group consider the possibility of only reporting effect sizes with statistically significant findings to avoid potential confusion.

Lane summarized next steps based on the discussion. She recommended that she work with Dvorak to provide examples of how effect sizes could be used in reporting to look within and across states, and across content areas. This will help COSDAM members better consider the utility of effect sizes in reporting and whether it can be meaningful to those who are not statisticians or psychometricians.

Peisch joined the call.

## Reaction to Achievement Levels Communications Plan

Lane introduced the next discussion as an opportunity to review plans for improving achievement levels communications. She noted there are two types of documents to be created: (1) brief communications documents targeted at specific subject areas and stakeholders and (2) a validity evidence document. She reminded attendees of the purpose for conducting these activities, and some of the recent decisions made by COSDAM.

Lane first inquired as to whether the group agreed with focusing the brief communications documents first for the subjects of reading and math and on the audiences of journalists and governors and their staff. Other groups and subject areas would follow. She and Wright recommend these target audiences saying they are likely to be strong users of NAEP data, and offer two very different perspectives to consider. Hanushek, Pope, and Peisch expressed agreement. Wright expressed that journalists will need solid information to produce news articles whereas governors and their staff will want to know how to use the information to inform key policy decisions. Peisch agreed and noted that it would be best to keep documents geared towards governors and their staff at one-page, and journalists may need a slightly longer document.

Lane next presented examples of how item map information and released exemplar items might be used in brief communications documents to help add clarity surrounding achievement levels. The example was taken from communications documents generated for the recent NAEP U.S. History and Civics release. All COSDAM members agreed with including this type of information in brief communications documents. Pope felt including this information was very powerful, and Wright noted the example questions provide great clarity surrounding what students can do.

Lane walked through the steps for developing the brief communications documents, and for obtaining internal and external feedback throughout the process. COSDAM members generally agreed with the approach. Wright noted it may be difficult to get past or current governors to participate in focus groups to obtain feedback, so it may be best to focus on their education staff. She noted there are Republican and Democrat governors' organizations that could be good sources. Pope expressed his hope that these communications efforts lead towards use of NAEP results in positive ways, and that they are not weaponized or politicized.

Lane next described the process for developing a validity argument document. The purpose of this document will be to provide information on appropriate interpretations and uses of achievement levels, highlighting the supporting validity evidence. Lane walked through the plans for developing and obtaining feedback, and expressed that measurement professionals should be added to the list of stakeholders from whom external feedback is sought. COSDAM members were in agreement with the approach outlined for completing the validity argument document.

The session ended with agreement to move forward with the achievement levels communications plans as presented.

The meeting adjourned at 11:13 am ET.

I certify the accuracy of these minutes.

A handwritten signature in cursive script that reads "Suzanne Lane".

Suzanne Lane, Chair

06/23/2023  
Date

# **National Assessment Governing Board Reporting and Dissemination Committee Report of May 8, 2023**

**Reporting and Dissemination Committee Members:** Vice Chair Marty West, Tyler Cramer, Anna King, Julia Rafal-Baer, Ron Reynolds, Mark White.

**Reporting and Dissemination Committee Members Absent:** Alberto Carvalho.

**National Assessment Governing Board Staff:** Becky Dvorak, Stephaan Harris, Laura LoGerfo, Lesley Muldoon, Munira Mwalimu, Sharyn Rosenberg, Elizabeth Schneider, Angela Scott.

**Institute of Education Sciences Staff:** Tammie Adams, David Geary, Tina Love.

**National Center for Education Statistics (NCES) Staff:** Brian Cramer, Peggy Carr, James Deaton, Enis Dogan, Vera Edwards, Eunice Greer, Dan McGrath, Emmanuel Sikali, William Tirre, Ebony Walton, Grady Wilburn, Angela Woodard.

**Other attendees:** American Institutes for Research (AIR): Brittany Boyd, Markus Broer, Cadelle Hemphill, Young Kim; CRP: Shamai Carter, David Hoff, Anthony Velez; Educational Testing Service (ETS): Marc Berger, Kadriye Ercikan, Robert Finnegan, Helena Jia; Hager Sharp: James Elias, Joanne Lim, Erik Robelen; The Hatcher Group: Melissa Meillor; Lerner Communications Group: Michelle Lerner, Nancy Zuckerbrod; Mathematica: John Deke, Jonathan Gellar, Brian Gill, Jennifer Starling; Optimal Solutions: Daniel Loew, Peter Simmons; Pearson: Joy Heitland.

The Reporting and Dissemination (R&D) Committee convened virtually via Zoom on Monday, May 8, nine days prior to the quarterly board meeting. Chair Alberto Carvalho could not attend; Vice Chair Marty West assumed chair responsibilities. West called the meeting to order at 12:32 pm EDT and outlined the agenda.

## **Debrief on the Release of the 2022 NAEP Civics and NAEP U.S. History Assessments**

West invited Laura LoGerfo, Assistant Director for Reporting and Analysis, to discuss the May 3 release of the 2022 NAEP Civics and NAEP U.S. History results. LoGerfo began by framing the context into which the Board released the data. News about Texas' choices on civics and history curricula primed the education sector to seek

information on student knowledge and skills in these subjects. The Board communications team — LoGerfo, Stephaan Harris, Michelle Lerner, and Nancy Zuckerbrod — also drafted a media advisory, notifying journalists that results would be released. In addition to journalists, advocates within the philanthropy space pointed to the upcoming results as critical evidence to support their calls to action.

Given this crowded landscape, the release event focused on reaching the nationwide audiences who could attend the event online. To promote the release event, four promotional videos which quizzed passersby on the National Mall about their U.S. history and civics knowledge from actual NAEP questions also generated interest in the results. These videos garnered 65 engagements and an average potential reach of 1,363 per post across all platforms.

The event followed the plan as approved by the Governing Board at the March 2023 quarterly meeting. Matt Stern of the Governing Board staff helped secure a room at the Capitol Visitors Center. The location fit the content and theme of the assessments perfectly, especially in that an alarm would sound periodically to prompt senators to vote. This served as a useful reminder of civic responsibility, representative democracy, and the republic at work.

Patrick Kelly, Governing Board member and grade 12 government teacher from South Carolina, opened the event with remarks explaining the vital importance of learning about government and history. Dr. Peggy Carr, commissioner of the National Center for Education Statistics (NCES), presented the results for the NAEP U.S. history assessment first, then sat at a table with Patrick Kelly to discuss those results and answer his questions.

Dr. Carr then returned to the podium and shared the NAEP Civics results, after which she addressed questions pre-recorded by directors, docents, students, and rangers at five locations of historic and civic significance across the country. Patrick Kelly teed up each question and, at the conclusion of the response to the fifth question, thanked Dr. Carr, the Capitol, and the audience.

Beyond the event itself, newspaper and television news programs interviewed Dr. Carr, Patrick Kelly, and Marty West. As with previous releases, the Board posted a media list along with a press release on the site where media access the embargoed results early to meet publication deadlines. This document lists the Board members available for interviews, which facilitates journalists' work. Kelly appeared on PBS NewsHour for an extensive discussion about the results, and West explained the results and trends on

the CBS Evening News. Both addressed questions and interpreted the meaning and import of the results for general audiences.

To facilitate accurate interpretations of the results and to convey what content the assessments cover, the Board communications staff generated materials that focused on specific NAEP assessment items. These helped to inform audiences that NAEP measures factual content, transcending political debates currently roiling the field.

More than 800 people registered to attend the release event virtually and 399 people joined on the day of the event. Several organizations that have engaged with NAEP content on the Board's social media channels since September re-emerged for the civics and U.S. history release, including Curriculum Matters, Knowledge Matters, 50CAN, and the U.S. Chamber of Commerce Foundation.

LoGerfo finished briefing the committee on the release event and related activities and encouraged questions. Ron Reynolds reported that the release was enjoyable to watch, especially the warm, insightful discussion between Dr. Carr and Patrick Kelly. Reynolds' fellow committee members concurred, corroborating that this approach to a release proved novel and engaging.

Several R&D committee members agreed that the emphasis on the actual content through the release materials should be replicated in the future. Reynolds inquired whether the Board developed a strategy to drive people to view the video of the release. Indeed, the Board's website will include an entire page devoted to the release materials, the recording of the event, and the promotional videos.

### **Release Plan for 2023 Long-Term Trend – Age 13: Discussion and ACTION**

LoGerfo next shifted attention to the upcoming release of the 2023 Long-Term Trend results for age 13. This release represents the final set of results from the suite of pandemic-era NAEP administrations. Similar to prior LTT releases, the 2023 LTT release offers only national data, from a reading test and a mathematics test of essentially basic skills, on trend lines that stretch for more than four decades, with limited contextual data.

The plan for the release event includes opening remarks by Marty West, the statistical presentation by Dr. Carr, and a panel discussion, moderated by West, among three teachers and three students in grades 9 and 10 about what supports and resources they need to accelerate learning. Rather than dwelling on how COVID affected the school experience two years ago, this conversation will stretch forward and seek

insights on what should come next. The students who participated in NAEP will finish their secondary schooling in three short years. What can be done now to help them?

The release will promote three primary messages, spotlighting NAEP's relevance and power, both for data from this specific release of LTT Age 13 and for data across the multiple NAEP assessments released in the last nine months. First, the Governing Board and NCES nimbly pivoted the assessment schedule and administration, respectively, to capitalize on an opportunity to collect and release important achievement data in a timely fashion. Second, the 2023 LTT measures students' knowledge and skills in content fundamental to success in high school and beyond, re-emphasizing how urgently everyone in education should be investing in learning acceleration. Third, for the last nine months, NAEP has provided invaluable data that stakeholders across the country use to galvanize efforts to accelerate learning, to find learning gaps that must be closed, and to invest fleeting funds in best practices. This is the last release of new assessment data from NCES for at least a year. Stepping back and out to synthesize lessons from across these assessments is a necessary endeavor for all in education.

LoGerfo sought feedback from the committee members on the plan. West cautioned that the Board's materials and messaging should distinguish Long-Term Trend (LTT) from Main NAEP but should not belabor the differences. He noted that people seem exhausted from the flood of new data in the last year, having become inured to dire assessment results. Orienting the discussion to the future, spotlighting teachers' and students' perspectives, and describing successful recovery efforts should prevent audiences from tuning out.

Julia Rafal-Baer agreed and suggested highlighting core recovery strategies, such as high-dose tutoring, mental health supports, and workforce development paths. Ron Reynolds advised demonstrating the potential utility of NAEP results to acceleration efforts but warned that the Board cannot appear to endorse any specific policy recommendations. West seconded this warning and refined his recommendation: the Board can elevate, not endorse, strategies shared by practitioners.

Mark White recommended reusing and recycling LTT language from the September release of age 9 LTT results. Rafal-Baer urged finding a space for the event that embraces pluralism, such as a charter school. Anna King extended that idea by recommending hosting the event beyond Washington, DC, to include more potential stakeholders. Rafal-Baer seconded that suggestion and warned not to burden release events with outsized expectations. The release event allows the media to center their stories on a date and location, but the Board should create space for dialogue outside the event and sustain conversation over a more extended time period.

With no further questions or suggestions, West called for the committee to move sending the release plan for the 2023 LTT results for age 13 to the full Board for action at the quarterly meeting in Los Angeles. This motion was approved unanimously.

### **Alternative Approaches to Issues of Statistical Significance**

The committee then turned to a presentation from Brian Gill and his colleagues at Mathematica. Gill and his team [penned a post](#) that took issue with NAEP reporting which presents non-significance in score comparisons as no change. To them, the absence of a statistically significant difference does not indicate no change and can obscure a change that is large enough to count as educationally meaningful.

Concerns about flukes in the data appropriately motivate checking score differences for significance; do calculated differences capture true differences in scores across time or between populations, or do the estimations of those differences simply reflect random quirks of a particular sample? Less technical audiences for NAEP data, e.g., policymakers, state education leaders, district administrators, parents, may not understand the statistics that underlie significance testing. Instead, these audiences default to trusting only the binary decision of significance or not, which, Gill argues, is insufficient to determine whether a result is real or meaningful. A non-significant result means only that *if* the true score did not change, the likelihood that a change that big would occur by chance alone is greater than 5%. Alternatively, a difference that is statistically significant at a 95% confidence level does **not** mean the result is 95% likely to be real.

Statistical analyses generate confidence intervals around estimates, which express how much uncertainty surrounds a given decline. Larger sample sizes result in narrower confidence intervals. In traditional statistical significance testing, if zero is within this confidence interval, the decline is not deemed significant, even though most of the interval shows an increase or decrease.

Bayesian methods are designed to avoid such misinterpretation, because rather than a binary signifier, Bayesian approaches assign a probability to a given change. The Bayesian approach “borrows” information from (1) jurisdictions at a single point in time; or (2) across subjects/grades within a given jurisdiction; or (3) across time points in a single jurisdiction. The amount of borrowing depends on the extent of confidence in the results, e.g., smaller samples mean borrowing more information.

Bayesian analysis considers confidence intervals for other estimated differences (e.g., grade 4 math in Detroit) and whether those intervals include or exclude zero. If most of



the interval for grade 8 math is below zero, then the Bayesian approach allows an interpretation of “X percent of that interval is below zero, so we are X percent sure of a decline.”

Gill explained this “borrowing” process through his team’s re-analysis of Detroit’s grade 8 NAEP reading scores released last October. The re-analyses are informed by other Detroit data and data from districts similar to Detroit, essentially running correlations across sites or assessments to determine how much adjusting is required to improve inferences based on assumptions about the underlying score distributions.

Gill and his colleagues summed their argument by acknowledging that what NCES presents rightly accounts for random flukes in the data but by recommending Bayesian analysis for dealing with random variation in a way that improves the information presented. Bayesian methods use all available data to account for flukes, downplay extreme values especially with small samples, avoid presenting results that can be easily misinterpreted, and provide results that lend themselves to easier interpretation.

While explaining the approach, Gill and his co-presenters addressed questions and concerns posed by both Governing Board members and by NCES staff. Dan McGrath, who leads assessment at NCES, did not favor adding excessive layers of analysis and interpretation on official statistics, suggesting that official statistics should be distinguished and kept separate from statistical analyses conducted by researchers. Enis Dogan from NCES cautioned that the assessment scales are not as vertical as assumed by Gill’s models and questioned why another district’s scores would bear any relevance to Detroit’s scores.

Gill replied that these issues really mount a challenge in communications. Analysts deploying Bayesian methods need to account for some similar districts, but not others, depending on the extent to which the districts are empirically related.

Gill’s colleagues admitted that these methods do not represent computationally difficult or time-consuming efforts, but that education rarely, if ever, uses Bayesian methods, despite its popularity in other research fields such as public health.

Dr. Carr remarked that this approach may be facile for the researcher, but opaque for the public, thus potentially leading to misunderstandings. Gill responded that the status quo of statistical significance fosters misunderstandings as well, given that thresholds for significance seem arbitrary and already incorporate assumptions of interpretation that the public may not grasp.

West intervened to note that NAEP sets the gold standard for national assessment, and the Board must defer to NCES' reporting needs to ensure adherence to federal statistical policy. With that, West concluded the meeting at 2:47 pm EDT, 17 minutes past its intended end time and thanked all for a rigorous and robust conversation.

I certify the accuracy of these minutes.

Alberto Carvalho

Alberto Carvalho, Chair

June 25, 2023

Date

# National Assessment Governing Board

## Nominations Committee

### Report of May 2, 2023

#### CLOSED SESSION

**Nominations Committee Members:** Reginald McGregor (Chair), Tyler Cramer, Viola Garcia, Suzanne Lane, Scott Marion, Alice Peisch, Ron Reynolds.

**Committee Member Absent:** Nardi Routten.

**Board Staff:** Lesley Muldoon (Executive Director), Elizabeth Schneider (Deputy Executive Director), Stephaan Harris, Munira Mwalimu, Tessa Regis.

Under the provisions of exemptions 2 and 6 of § 552b (c) of Title 5 U.S.C., the Nominations Committee met in closed session on Tuesday, May 2, 2023, from 5:00 p.m. to 6:30 p.m. Eastern Time (ET) to discuss ongoing work.

Chair Reginald McGregor welcomed members and called the meeting to order at 5:00 p.m. He provided a preview of the following agenda topics:

- Update on Nominations for Board Terms beginning October 1, 2023
- Review and Discussion of the Nominations Processes
- Proposed Communications Plan
- Review of Policies and Procedures Manual
- 2024 Nominations Outreach Campaign

Elizabeth Schneider provided an update on the finalists for the 2023 term of office. She briefed the committee on a meeting she and Lesley Muldoon had with Department officials to alert them to the historic number of vacancies for the 2023 term. A meeting will soon be convened with the Department's White House Liaison.

Schneider then previewed the vacancies for the 2024 nominations cycle, noting that there are vacancies in six categories:

- General Public Representative (Parent Leader)
- Local School Board Member
- Non-Public School Administrator
- State Legislator (D)
- State Legislator (R)
- Testing and Measurement Expert

Schneider also noted that five current members are eligible for reappointment; Alice Peisch, state legislator (Democrat), has served two full terms and is not eligible for reappointment.

Next, McGregor requested feedback from members on the completed 2023 nominations cycle and suggestions for improving the 2024 campaign. Members discussed the communication protocols with applicants, eligibility clarification, and the need for expanding outreach for the 2024 Nominations cycle.

Finally, Stephaan Harris shared plans for outreach in the 2024 cycle and emphasized the importance of engaging the Board members and conducting outreach to increase the racial and ethnic diversity of the applicant pool.

McGregor thanked members and adjourned the meeting at 6:30 p.m.

I certify the accuracy of these minutes.

  
Reginald McGregor, Chair

May 12, 2023  
Date