Assessment Development Committee

May 17, 2023
10:30 am – 4:45 pm PT
Golden State Ballroom II

AGENDA

10:30 – 11:00 am	**Discussion of Consultant Papers on 2030 NAEP Writing Assessment Framework** *Patrick Kelly, Chair*	Attachment A
11:00 – 11:30 am	**Potential Updates to Assessment Framework Development Policy** *Christine Cunningham, Vice Chair*	Attachment B
11:30 – 12:30	**Policy Guidance on Proposed 2028 NAEP Science Assessment Framework** *Patrick Kelly* *Mark Loveland, WestEd* *Taunya Nesin, WestEd*	*See Plenary Tab*
12:30 – 1:45 pm	**Lunch Break**	
1:45 – 2:15 pm	**Budget Considerations for Framework Updates (CLOSED)** *Patrick Kelly*	
2:15 – 4:45 pm	**Review of 2024 Pilot Mathematics and Reading Items at Grades 4 and 8 (CLOSED)** *Patrick Kelly* *Christine Cunningham*	*Secure materials sent separately*
Information Item	**Item Review Schedule**	Attachment C

Discussion of Consultant Papers on the
2030 NAEP Writing Assessment Framework

According to the NAEP Assessment Schedule, the NAEP writing assessment will next be administered in 2030 and updates to the framework will be considered for this administration.

The current NAEP Writing Assessment Framework was adopted in 2007 for implementation in 2011. The Board made a policy decision at the time to begin new trend lines without attempting to perform bridge studies to determine the feasibility of connecting results based on the previous framework. The current framework focuses on "writing on computer," replacing the previous framework which focused on writing by hand; the mode of administration in the current framework is not incidental but is conceptualized as being a central part of the construct.

The framework assesses three communicative purposes at grades 4, 8, and 12: *To Persuade, To Explain,* and *To Convey Experience, Real or Imagined*. Each sampled student receives two 30-minute computer-based writing tasks and has access to software similar to common word-processing programs. The framework calls for a specific audience to be stated or clearly implied.

In 2011, the NAEP writing assessment was administered at the national level at grades 8 and 12; results from that administration can be found at: https://www.nationsreportcard.gov/writing_2011/. In 2017, the NAEP writing assessment was administered at the national level at grades 4 and 8, but the results were not able to be reported due to technical concerns related to changes in the device and platform used to administer the assessment; more information is available at: https://nces.ed.gov/nationsreportcard/writing/2017writing.aspx. The technical issues encountered in 2017 make it very unlikely trend can be maintained with the 2011 results in the future, regardless of whether or not a new framework is adopted.

In accordance with the Board policy on Assessment Framework Development, the first step in the process of updating a framework is to seek public comment on whether and how the existing framework should be changed. Following the ADC discussion at the November 2022 quarterly Board meeting, an open call for initial public comment on the current NAEP Writing Framework was conducted from November 29, 2022 – January 25, 2023. Commenters were asked to address three questions:

- Whether the NAEP Writing Assessment Framework needs to be updated
- If the framework needs to be updated, why a revision is needed
- What a revision to the framework should include

Twenty-one submissions were received from a variety of individuals, groups of individuals, and organizations. In addition, Board staff sought input from the National Center for Education Statistics (NCES) on operational issues and challenges associated with the current framework and assessment; a memo was submitted by NCES to summarize their feedback. The public comment summary and NCES memo were shared with ADC as part of the March meeting materials and are attached again here (the raw individual comments can be found in the March ADC materials if anyone is interested).

Next, Board staff commissioned short papers[1] from the following writing experts to inform additional discussion at the upcoming May ADC meeting:

- *Elyse Eidman-Aadahl*, Executive Director, National Writing Project
- *Carol Jago,* Independent Consultant and former Governing Board member/ADC Chair
- *Sandra Moumoutjis,* Executive Director, Learning Innovation Network at Building 21
- *Peggy O'Neill*, Professor of Writing, Loyola University of Maryland
- *Tonya Perry*, Vice President, National Council of Teachers of English[2]

During this session, ADC members will discuss the consultant papers and their implications for whether and how to proceed with updating the 2030 NAEP Writing Assessment Framework.

Consultant Biographies

Elyse Eidman-Aadahl is currently the Executive Director of the National Writing Project (NWP). Prior to assuming that position, she was Director of National Programs where she developed NWP's signature national programs, action research networks, and scoring systems. Formerly a high school English and Journalism teacher, a teacher educator and university professor, Dr. Eidman-Aadahl's research interests focus on the impact of digital tools and internet connectivity on writing and the teaching of writing, writing and civic engagement, and how teachers learn and reason about practice. Since 1996, she has served on numerous NAEP committees for writing, including the Framework and Standing Committees for the 2011 NAEP Writing Framework. A recipient of several awards for teaching and scholarship, she has a Ph.D. in Curriculum Theory from the University of Maryland and sits on the Board of the Center for the Collaborative Classroom.

Carol Jago has taught English in middle and high school in public schools for 32 years and is associate director of the California Reading and Literature Project at UCLA. She served as president of the National Council of Teachers of English and as chair of the College Board's English Academic Advisory committee. She has published many books with Heinemann including *The Book in Question: Why and How Reading Is in Crisis.* She is also author of *With Rigor for All* and *Cohesive Writing: Why Concept Is Not Enough* and published books on contemporary multicultural authors for NCTE. Ms. Jago was awarded the International Literacy Association's Adolescent Literacy Thought Leader Award and the CEL Exemplary Leadership Award. She has received a Lifetime Achievement Award from the California Association of Teachers of English. Ms. Jago has served on the National Assessment Governing Board and currently serves on the International Literacy Association's Board of Directors. She is also the recipient of the National Council of Teachers of English Squire Award given to honor an individual who has had a transforming influence and has made a lasting intellectual contribution to the profession.

Sandra Moumoutjis is the Executive Director of Building 21's Learning Innovation Network which is designed to grow and support a community of schools and districts across the country as

---

[1] The consultant papers were produced under Governing Board contract number 919995921F0002 to Manhattan Strategy Group.

[2] Tonya Perry's paper was not available yet at the time the Board materials were finalized and will be sent separately

they transition to personalized and competency-based education. Through professional development and intensive coaching, Ms. Moumoutjis supports schools and districts in all aspects of the change management process. Ms. Moumoutjis is the co-designer of Building 21's Competency Framework for students, teachers, and leaders as well as the studio instructional model. She has written a blog series for the Aurora Institute outlining her vision for the possibilities of competency-based education to meet the unique needs of every learner. Prior to working for Building 21, Ms. Moumoutjis was a teacher, K-12 reading specialist, literacy coach, and educational consultant in districts across the country.

Peggy O'Neill, PhD, is a Professor of Writing at Loyola University Maryland. She entered the education profession as a middle school English teacher and has been a college professor for 25 years. During her career, she has served as the director of composition, department chair, and associate dean. The primary focus of her research and scholarship is writing pedagogy and assessment. She is the co-author or editor of seven books, including *Assessing Writing to Support Learning: Turning Accountability Inside Out* (2023), *Reframing Writing Assessment to Improve Teaching and Learning* (2010), and *The Guide to College Writing Assessment* (2009). Her research has also appeared in many peer-reviewed journals, including the *Journal of Writing Assessment*, *Assessing Writing*, and *College Composition and Communication,* as well as peer-reviewed edited collections. She has served on assessment-related committees for professional organizations such as the National Council of Teachers of English and the Council of Writing Program Administrators and as a consultant with a variety of organizations.

Tonya Perry is the Vice Provost and Full Professor at Miles College, an Historically Black College and University (HBCU) in Fairfield, Alabama. She is currently serving as the vice president of the National Council Teachers of English and an Executive Board member for the National Writing Project. As a writing project director, she specializes in working with students and teachers in writing instruction, creating curriculum that is responsive to the needs of students' writing development. Her writing project also has served as a writing assessment site, reviewing and studying student writing over several years of development, particularly the writing of Black and Brown students. Dr. Perry also has served on the National Board for Professional Teaching Standards (NBPTS) Global DEI Committee, developing guidelines and standards of practice for the next generation of the National Board standards. Her work is informed by her years as a teacher, teacher educator, professor, community-engaged scholar, and national consultant.

To: National Assessment Governing Board members and associated staff
From: Elyse Eidman-Aadahl, Executive Director
Date: April, 2023

I have been associated with NAEP in some form since the 1987-88 assessment. I was part of the Framework and Standing Committees for the [Writing Framework for the 2011 NAEP](#), and regularly consult with scholars, district leaders, and foundations regarding NAEP findings as Executive Director of the National Writing Project (NWP). Over that 30-year period, NAEP has led the field in pioneering and refining direct assessment of writing, including prompt development and testing, and primary/holistic/modified assessment practices. By the time of the Common Core State Standards (CCSS), which established broadly adopted anchor standards in writing, many of the key features of large-scale writing assessment pioneered by NAEP were firmly established in the profession and reflected in state assessment programs. This proud tradition of innovation and leadership, however, is at an inflection point. In this letter, I begin with an assessment and recommendation about the current Framework, then I discuss five issues that I believe the National Assessment Governing Board might consider as you deliberate about the future of NAEP Writing for 2030 and beyond.

The 2011 Framework for Writing is a strong framework.
Revisiting the 2011 Framework now, more than a decade after it was developed, shows it to be robust, relevant, even groundbreaking in its definition of writing as a purposeful act of communication embedded in a context which is both rhetorical (audience, purpose) and materials (digital tools, limitations of context). This basic definition of the act of writing, including its complex social and multifaceted nature, reflects current scholarship and continues to be relevant today. In addition, the framework strongly corrected for elements of previous frameworks that had become outdated. These corrections include:

1. the expansion of the purposes of writing to include those relevant to the workplace and to the writers' personal life in addition to academic purposes;
2. the shift from an emphasis on the "modes of discourse" and specific forms for writer-to audience purposes for writing: to persuade, in order to change the reader's point of view or affect the reader's action; to explain, in order to expand the reader's understanding; and to convey experience, real or imagined; and finally,
3. the recognition that digital tools and internet-connected reading/writing practices have profoundly reshaped the core experience and uses of writing itself.

These improvements from previous frameworks position the 2011 NAEP Writing Framework as a still-relevant foundation for assessments in the near term. At the same time, current generation assessments and developments in the digital tools and internet technologies raise significant questions that a future framework will need to address. I believe a prudent path would be to administer a final assessment to fully conclude the 2011 NAEP Writing Framework followed by the creation of a revised framework for 2030 to address challenges that were not met in the 2022 Framework and to establish new trendlines.

Why should NAEP conduct a final administration of 2011?

As the nation's only national assessment program, NAEP provides essential data about student performance in writing. Unfortunately, issues with the administration of assessments in 2017 mean that the nation has not had a full set of national data on student performance in writing since the results from the 2011 administration were released in 2012. The decade since that first administration under the 2011 NAEP Writing Framework was significant for curriculum changes sparked by CCSS as well as the broad adoption of computer technology in schools and homes. Although many states have data on writing from state assessments, including data from assessment consortia that allow for comparisons by districts and states, the absence of an anchoring assessment is problematic for anyone concerned with national level policy: researchers, foundations, curriculum developers, professional development providers, and for the writing community itself. I have frequent conversations with national providers and district leaders who worry about the lack of national data during this important reform period in K-12 writing.

The length of time necessary to develop a new framework, pilot, and then administer an assessment extends the wait for new data. The nation would, then, be missing key data to allow us to understand the impact, if any, of curricular reforms of the 2010s— the very reforms that were lauded as "the return of writing". My overall recommendation, therefore, is to conduct one final assessment under the 2011 NAEP Writing Framework. This would need to be done with an eye toward addressing the problems that affected the 2017 assessment, of course, but the solving of those technical problems could be understood as a step toward reimagining the testing context with a view to administration of a new framework to follow.

Why should NAEP develop a new framework for writing?

The 2011 NAEP Writing Framework was groundbreaking in its recognition of the impact of digital and internet-connected technology on writing, both in the academy and in the workplace. "The digital" is more than an administrative efficiency for the assessment or analysis tool after the fact. Digital composing tools are part of the core construct of the act of writing and affect the construct in a more direct way than in other assessments.

The impact of technology has only increased since the 2011 NAEP Writing Framework was developed. Writers compose on a wide range of devices, participate in social media and multimodal platforms, conduct search via the internet almost exclusively, and assess

credibility of sources in the context of internet circulation and information pollution. States and districts are developing standards for information and media literacy and creating capstone projects that require multimodal composition. Although the 2011 Framework called for writing to be produced in a word processing environment as a first step, the writing produced in the two 30-minute blocks of the current assessment will soon be seen as inadequate to speak to curricular expectations and conditions in schools and the workplace.

But even more significantly, the recent public release of generative AI such as ChatGPT has taken the discussion of technology and composing to a new level unforeseen by the 2011 NAEP Writing Framework because it so directly affects the core construct. ChatGPT has already led to concerns about the "standard school essay" as well as a general diminution of the kinds of writing collected in large-scale assessments. The public sees that large language models can produce similar, even superior results on these generic tasks and may come to see performance that samples this kind of generalized, quickly produced writing as uninteresting for assessment purposes. If the public comes to value new purposes for writing—such as writing for learning or the evaluation of evidence credibility— these purposes will require a significant reframing of the scholarship informing a new framework and a rethinking of the technical properties of prompts and administration. The announced timing of such a framework for 2030 provides a sufficient window to allow some of these issues to settle into place. With a final administration under the 2011 NAEP Writing Framework, a brief delay would be possible without further extending the nation's wait for data.

What key issues should be addressed in a new 2030 NAEP Writing Framework?

I believe that a 2030 NAEP Writing Framework will need to be a second "groundbreaking" framework, one that in effect resets the trendlines and models that characterize the current NAEP Writing assessment and, perhaps, push up against policy expectations that surround NAEP in general. This would return NAEP to its position as an innovative leader in assessment, one that has the freedom to innovate because there are no stakes for the individual test-taker.

Key issues that a new framework will need to address include:
- Limitations of the administration context
- Relationship to high-quality state assessments
- Relationship between NAEP prompts and use of source material
- Drawing writing measures from new generation reading assessments
- Employing machine technology for greater descriptive power

Limitations of the administrative context: All assessments need to take into account the limitations of the assessment context. For NAEP, limits on the time and setting for overall administration mean that the writing assessment is typically 60 minutes per student. To employ the balanced incomplete block design (BIB spiraling) used for scaling, students

must be asked to complete two equivalently timed prompts, yielding 25 to 30 minutes per task. During the 1990s, NAEP included 50-minute tasks, but the results were not scalable.

The 30-minute restriction places very tight limits on what can be assessed. Writers have insufficient time to examine complex stimuli, read sources or view informational media, or to plan and draft more than a few paragraphs. We already see that longer periods of time and more complex, source-driven stimuli are used in many state assessments, and it is likely that growing questions about young people's performance in a digital age will focus on these more complex tasks and stimuli.

If the 30-minute limit is maintained, NAEP will not be able to provide information about more sophisticated performance targets. Already it was the case that aspects of performance proposed in the 2011 NAEP Writing Framework could not be accommodated in 30-minute blocks. A new framework must address this problem head-on and come to a resolution that positions NAEP usefully in the context of an assessment system that looks quite different in 2023 than in the mid-2000's when the 2011 NAEP Writing Framework was being developed.

Relation to state assessments: In contrast to the 2000s, NAEP now participates in a context of well-developed state assessments that were originally inspired by the CCSS assessment consortia. Reports such as the Expert Judgment Comparison produced by the NAEP Validity Studies Panel have looked in detail at the differences and similarities between NAEP Writing and current generation state assessments. Judged simply on elements such as length of time, rigor of stimuli, etc., many state assessments set a higher mark than NAEP at assessing some of the performances recommended in the 2011 NAEP Writing Framework. NAEP, however, provides important services to states that cannot be provided by their own assessments: namely, 1) a nationally representative and comparative sample, and 2) a "no stakes" environment for students that allows for a level of experimentation and innovation that would be unwise in local high stakes assessments. Within this new assessment context, a new NAEP Writing Framework could explicitly consider the question of what limited performance measures would be useful to states as comparative national measures intersecting with the measures they already collect. Attention to sampling frames that would allow for explicitly equating performance might be a useful provision as would the involvement of states in setting research questions for surveys and supplemental studies.

Relationship between NAEP prompts and the use of source material: According to the Expert Judgment Comparison review, all current generation state assessments that assess writing have at least some portion devoted to writing with sources, and writing with sources predominates in many. In our own writing assessment programs at NWP, we have moved exclusively to writing with sources in order to meet the interests and needs of our participating districts and research partners. NAEP is the only assessment reviewed in the report where 100% of the prompts are "writing without sources" prompts (Expert Judgment Comparison, p. 13).

Significantly, the move toward writing with sources followed from increased attention to argumentative writing and the ability of students to assess and use evidence for claims. The move toward argument writing was central to CCSS and has remained consistent across states even in states where CCSS was not adopted or is not relevant. Writing with sources is also an interest of the higher education community and, increasingly, the workplace community. The use of sources and the ability of students to understand and use data displays, infographics, and primary source evidence are relevant to educators across the curriculum as well. In short, writing with sources is a high value skill.

Writing prompts that require the use of sources, in contrast to prompts where additional stimuli are included simply to "spark writing" but are not required to be used in the writing, obviously push the limits of the testing context. They also introduce additional challenges that a new framework would need to address: 1) what sources should be included at what level of difficulty? 2) how should accuracy in use of sources be judged and to what level of significance? 3) how should attention to understanding of sources be balanced with attention to the effective use of the sources in written performance? and 4) how do we choose sources that work across linguistic, cultural, and social demographics in American schools?

These are very complex determinations irrespective of questions about the length of the testing context.

Drawing writing measures from new generation reading assessments: Frequently, the suggestion is made that writing measures could be drawn from reading assessments where students are required to write extended responses to reading. Such an approach would seem to have the potential to create a more suitable assessment context by "borrowing" minutes from the reading assessment and making NAEP Writing more like current generation state assessments.

This is also a complex issue for a new framework. The writing community has generally rejected short, constructed response passages as sufficient to assess anything other than the most basic skills of language facility and use of conventions. There is insufficient space to demonstrate the rhetorical judgment and mastery of content that is central to improvement in writing. In addition, the habits of "schoolish" response to reading often constrain writers, encouraging them to reproduce habits more appropriate to transactional writing than to fuller composing. Deriving writing measures from reading assessments risks underplaying the complexity of the reading/writing interaction and positioning writing as a mere handmaiden to reading.

State assessments address this by creating new "ELA" measures where the stimulus is constructed to build off of reading, yet still invite a rhetorically situated response. This new type of item is not quite a reading item and not quite an open writing item. Depending on the construction of a reading assessment, it may or may not be able to "borrow minutes"

from an assessment with other purposes. It is also an item that works more effectively at the 8th and 12th grade level than at the 4th grade level. Previously, NAEP has presented the grade level assessments as a continuing arc of performance, but in this case the problems manifest more strongly at the 4th grade level. In summary, it is an intriguing possibility, but one that would require the full attention of a framework committee and, perhaps, a radically different approach.

Employing machine analysis for greater descriptive power: Finally, the 2011 NAEP Writing Framework nodded toward the potential for the rich analysis of student writing that is possible when digital texts are collected. NAEP was able to provide some reports in administrations under the 2011 Writing Framework, including an analysis of keystroke data. But the potential for framing the assessment around a significant brief for descriptive data was never fully explored as achievement level data was the focus. Now, however, with much more extensive achievement information coming from state, district, and school-based assessments, the balance might be reconsidered. New technologies allow for a much fuller description of the corpora of student work, providing detailed information to the nation about what young people are able to do and what they struggle with. This rich descriptive and analytic information would go much further in informing curriculum and instruction than more simple achievement level descriptions.

In summary, drawing mostly from my interactions with the writing community, discussions with local and state leaders, and my own knowledge and review of the 2011 NAEP Writing Framework, I urge the Governing Board to consider conducting one final assessment of under the auspices of the 2011 NAEP Writing Framework, then planning for a robust new framework for 2030.

The NAEP Writing Framework Revisited

Prepared for the National Assessment Governing Board
By Carol Jago
March 2023

As the National Assessment Governing Board, and particularly the Assessment Development Committee, ponder whether or not to begin the process of revising the 2017 NAEP Writing Framework, I offer three observations for consideration:

1. The underlying principles of sound writing assessment have not changed in the intervening years. The following statement from the 2017 Executive Summary could have been written yesterday:

    > "The use of written language has become a critical component of the daily lives of millions of Americans. This is in part because, as technology continues to alter societies and cultures, it has fostered and supported an unprecedented expansion of human communication. To write in this world is to engage in a millennia-old act that is reinventing and regenerating itself in the modern age.
    >
    > The impact of communications technologies has changed the way people write and the kinds of writing they do. Writing in the 21st century is defined by its frequency and efficiency, and modern writers must express ideas in ways that enable them to communicate effectively to many audiences. It is clear that the ability to use written language to communicate with others—and the corresponding need for effective writing instruction and assessment—is more relevant than ever."

    There is no need for a new Writing Framework. Instead, NAEP needs to address the challenges that implementing the 2017 Writing Framework posed for both students and item writers.

2. The disturbance in the field that has the writing community scrambling to catch up is the emergence of artificial intelligence (AI) in the form of ChatGPT. Software capable of writing sophisticated responses to prompts has teachers, scholars, the entire composition community engaged in heated conversations regarding how this development will affect writing instruction and assessment. Every writing project institute, every professional development session this summer, will be exploring the impact AI is likely to have on student writing.

    The National Assessment Governing Board should wait for and/or possibly support research that better defines the issues surrounding the use of ChatGPT and other AI

tools. At the moment, school and district policies regarding its use range from embracing the tool to forbidding it from campus. This is not the moment to revise the NAEP Writing Framework.

3. As a former Governing Board member, I understand the need to set priorities based upon available funding. Engaging in a full framework revision process is both expensive and enormously time-consuming for staff. Given how rarely the NAEP writing assessment has been administered, a more prudent use of funds might be to have a technical advisory committee make recommendations based upon lessons learned from past administrations and to implement a writing assessment as soon as feasible.

Does the 2017 NAEP Writing Framework continue to reflect current trends in the field of writing including state writing standards and assessments?

To a large degree, yes. The first iteration of this framework in 2011 garnered headlines for the shift to computer-based assessment. The speed of technological change has increased over the ensuing years and has made this shift ever more important (and at the same time ever more challenging to implement). It was interesting to note that the concern about 4th graders' keyboarding skills that appeared in the public comments in 2023 is exactly the concern of several Framework Committee members in 2007 when the original framework was being written. I served as a member of that Writing Framework planning committee.

The 2011/2017 framework broke new ground with its designation of communicative writing purposes, a departure from identifying writing "types."

- To persuade
- To explain
- To convey experience, real or imagined

These purposes for writing rather were re-envisioned in the Common Core State Standards (CCSS). In that document which influenced curriculum, instructional materials, and practice nationwide (whether or not the name CCSS was changed) the communicative purposes for writing were combined with the more familiar text types:

Common Core Text Types and Purposes

- Write arguments to support claims in an analysis of substantive topics or texts.
- Write informative/explanatory texts to examine and convey complex ideas and information.
- Write narratives to develop real or imagined experiences or events.

State assessments of writing commonly reflect these three types of writing resulting in teaching that aims to replicate particular forms of writing.

Some writing scholars and teachers are keen to expand the field of writing to include multi-media: video, images, Tik Tok, podcasts as contemporary vehicles for expression and communication. This expansion would, I believe, entail the creation of a whole new NAEP assessment, possibly in Communication, wherein actual writing played a supporting rather than central role.

The introduction to the 2017 NAEP Writing Framework asserts:

> "The ability to write well is essential to the economic success of the nation. Americans in the 21st century need to be able to communicate in a variety of forms and mediums, create texts under the constraints of time, and play a productive role in an economy that increasingly values knowledge and information. The pace of written communication in today's environment—the velocity of writing— reflects the transition to an information-based economy built on speed, efficiency, and complexity" (1).

These conditions are as true today as they were when these words were written.

Are there disturbances in the field of writing that require major revision?

Yes.

On November 30, 2022 an artificial intelligence chatbot opened to the public. OpenAI invited all interested parties to test how ChatGPT could reply to a given prompt across a range of disciplines with detailed, well-written responses. While factual accuracy is often uneven, the chatbot's responses are free of mechanical errors and appear articulate. Teachers experimenting with ChatGPT on a student assignment found its generated text was similar to what a good student would deliver and began to worry if writing instruction as we have known it would soon become a thing of the past.

Fearing widespread plagiarism and that access to ChatGPT would be detrimental to students, New York City Public Schools immediately banned its use citing, "negative impacts on student learning, and concerns regarding the safety and accuracy of content," adding that "the tool may be able to provide quick and easy answers to questions, [but] does not build critical-thinking and problem-solving skills, which are essential for academic and lifelong success." Los Angeles Unified School District temporarily blocked access to ChatGPT and the OpenAI website in December, 2022 "to protect academic dishonesty, while a risk/benefit assessment is conducted." Baltimore City Public Schools followed suit.

Some schools quickly mobilized to write policy statements regarding the use of ChatGPT on their campuses. A private K-12 school in California adopted the following position:

> "Aligned with our English Department philosophy statement, the English department deeply values individual thought and critical thinking. Consequently, use of an AI text generator is permitted only if the assignment explicitly calls for it, requires it, or allows it. When permitted, students must follow specific assignment guidelines to disclose which text comes from AI. Content generated by an Artificial Intelligence third-party services or sites, including AI-generated content, without proper attribution, authorization, and expressed permission in the assignment, is considered a form of plagiarism."

Others in the field have embraced the possibilities the software holds for assisting students as they learn to write. On February 28, 2023, Dr. Mark Glanville, head of assessment principles and practice for the International Baccalaureate Program published the following in *The Times*:

> Those of us who work in the schools or exam sector should not be terrified by ChatGPT and the rise of AI software – we should be excited. We should embrace it as an extraordinary opportunity.
>
> Contrary to some stark warnings, it is not the end of exams, nor even a huge threat to coursework, but it does bring into very sharp focus the impact that artificial intelligence software that can write sophisticated responses could have on the way we think about teaching, learning and assessment.
>
> We should not think of this extraordinary new technology as a threat. Like spell-checkers, translation software and calculators, we must accept that it is going to become part of our everyday lives, and so we must adapt and transform education so students can use these new AI tools ethically and effectively.
>
> The International Baccalaureate, where I am head of assessment principles and practice, has decided it is not "banning" the use of ChatGPT or any similar AI software as has been seen elsewhere; this is the wrong way to deal with innovation.
>
> Students will, however, need to be made aware that we do not regard work written by such tools to be their own. To submit AI-generated work as their own is an act of academic misconduct and would have consequences. But that is not the same as banning its use.
>
> In truth, many of the issues thrown up by Chat GPT are extensions or variations of current issues that the IB is familiar with managing, even if these technologies are significantly different in terms of speed, ease of access and scale.

With such widely diverging points of view on the use of this new technology, the wisest course for the Governing Board to take is to wait until the dust settles and there is greater consensus

as to the issues involved. NAEP frameworks are built upon the foundation of rigorous research. At the moment, too many questions regarding the implications of artificial intelligence and student writing remain unanswered.

Possible courses of action

- Postpone any major revision of the NAEP Writing Framework until research studies can be conducted regarding the impact of AI on the teaching and assessment of writing.

- Charge a technical advisory panel to address the implementation challenges that the 2017 Writing Framework has posed for both item writers and students taking the assessment. The panel could make recommendations to the Governing Board regarding how the current framework can be adapted to address the issue of time allotted for writing tasks (30 minutes) which limits:

  o the stimuli materials that item writers can include
  o the variety of audiences that students can be asked to write to
  o the nature of the realistic persuasive scenarios that can be created

  Would it be possible to have 8th and 12th graders write in one 60-minute block rather than two 30-minute blocks? This adjustment would make room for the kind of scenario-based prompts that more closely match the vision of the 2017 Writing Framework.

- Follow the development of the use of automated scoring of student writing by other large-scale assessments (state tests, ACT, AP exams, etc.). If viable, the use of AI could dramatically lower the cost of administering the NAEP Writing assessment and thus allow for more frequent administrations of the exam.

- Consider the possibility of using the longer constructed responses from other assessments — for example Reading, Science, TEL, or U.S. History — as a measure of students' writing proficiency. Obviously, there would be many moving pieces to this plan, but it could obviate the need for a separate NAEP writing assessment and place the writing itself within more authentic circumstances. Students would be genuinely writing to persuade, to inform, and to convey experience.

Considerations for Updates to the NAEP Writing Assessment Framework
For Administration in 2030
Sandra Moumoutjis
Executive Director, Learning Innovation Network,
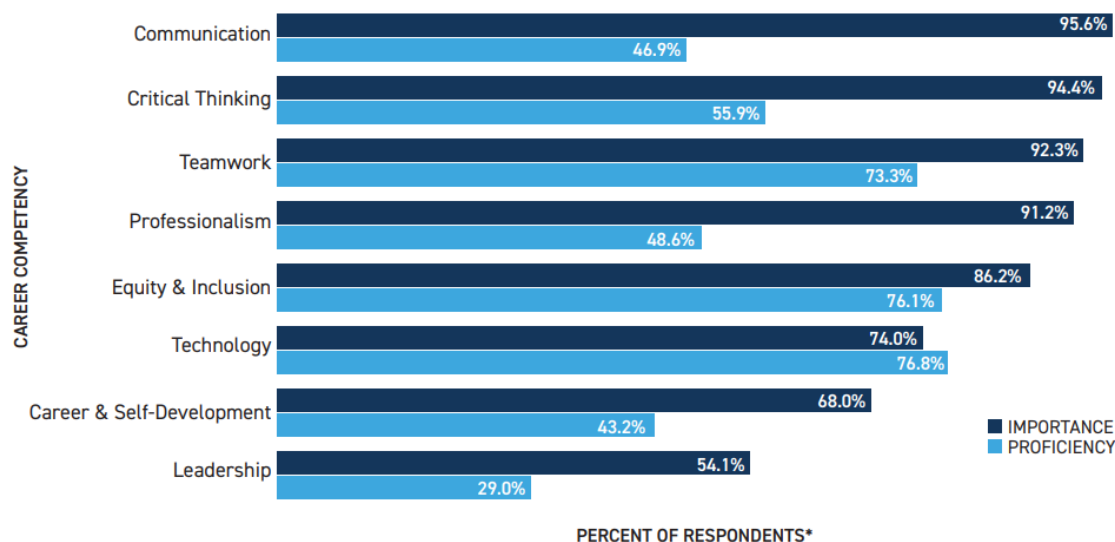Building 21
April 19, 2023

Introduction

The purpose of this paper is to determine whether and how the current NAEP Writing Assessment Framework should be updated for administration in 2030 and beyond. This paper will provide the National Assessment Governing Board with background on the current state of K-12 writing standards, learning, and assessment and specific recommendations for updating the NAEP Writing Framework.

Writing is an essential part of daily life. People write for a variety of reasons — to record information, to communicate with others, to tell stories and preserve culture, to process information, emotions, and events, to inform others, to entertain, to keep track of daily events or to chronicle events, travels, and experiences, to create imaginary worlds, to express feelings, and to entertain others (Graham, 2019). We tweet, we text, we Snapchat, we DM, we email — the various types of technological communication is increasing the need for everyone to be able to write effectively for many purposes (Slavin et al, 2019). The ability to express ideas in writing is one of the most important of all skills and is necessary for college, careers, and civic engagement.

According to the National Association of Colleges and Employers Job Outlook 2023, fifty percent of respondents indicated that written communication was an essential skill they were looking for on an applicant's resume. When considering the eight career readiness competencies, employers rate communication the most critical but, as the image below illustrates, it also is rated as having the largest gap in proficiency (NACE, 2022).

FIGURE 46: IMPORTANCE VS. PROFICIENCY ON CAREER READINESS COMPETENCIES, BY PERCENT OF RESPONDENTS

Career Competency	Importance	Proficiency
Communication	95.6%	46.9%
Critical Thinking	94.4%	55.9%
Teamwork	92.3%	73.3%
Professionalism	91.2%	48.6%
Equity & Inclusion	86.2%	76.1%
Technology	74.0%	76.8%
Career & Self-Development	68.0%	43.2%
Leadership	54.1%	29.0%

PERCENT OF RESPONDENTS*

*The percentages corresponding to "importance" represent, among all responding employers, the percentage that, on a five-point scale, indicated that the respective competency was either "very important" (4) or "extremely important" (5) for college graduates to have to enter their workforce. The percentages corresponding to "proficiency" represent, among all responding employers, the percentage that, on a five-point scale, rated recent graduates either "very" (4) or "extremely" (5) proficient in the respective competency.

A 2004 study by the National Commission on Writing, a panel established by the College Board, concluded that around a third of employees in the nation's blue-chip companies were poor writers and that as much as $3.1 billion was being spent by businesses annually to provide remedial writing training (Dillion, 2005). National writing assessment data supports the lack of proficiency the nation's students are demonstrating in the area of writing. Only twenty-seven percent of students in grades 8 and 12 performed at or above the Proficient level on the NAEP writing assessment in 2011 (National Center for Educational Statistics, 2012). In 2003, the National Commission on Writing issued a report labeling writing a neglected skill in American schools (National Commission on Writing, 2003).

Yet despite the abundance of evidence that we are failing to teach our students to be proficient writers, there is a lack of research about what needs to be done to improve the quality of our students' writing. According to Robert Slavin, director of the Center for Research and Reform in Education at the Johns Hopkins School of Education, "What's very odd about writing is how small the research base is. There's remarkably very little high-quality evidence of what works in writing" (Barshay & Barshay, 2019). Slavin and his Johns Hopkins colleagues, along with researchers in Belgium and the United Kingdom, were only able to find 14 studies that met their standards, but quickly found 69 studies just on teaching reading to high school students (Barshay & Barshay, 2019). Slavin also notes that a remarkable proportion of all research and reviews of research has been carried out by Steven Graham and Karen Harris, both professors at Arizona State University (Slavin et al, 2019).

What does the research tell us about the current state of writing instruction?

Below is a summary of the relevant findings related to the problems with writing instruction:

- While there are many factors that influence children's development as writers, including poverty, genetics, and biological functioning, many children do not receive the writing instruction at school that they deserve or need. Writing and writing instruction in most classrooms are inadequate. These findings were generally universal, applying across countries and grades (Graham, 2019)
- A majority of teachers do not devote enough time to teaching writing. At both the elementary and the secondary level, the typical teacher devotes much less than 1 hour a day to teaching writing (Graham, 2019).
- Students in a typical class do not write frequently (Graham, 2019). .
- There was a notable absence of the use of digital tools for writing (Graham, 2019).
- The number of students on a teacher's roster is a contributing factor in how writing is taught. As NCOW (2003) noted, it becomes increasingly difficult to provide writing instruction responsive to students' needs as the number of students in a classroom increases. Most middle and high school teachers have student loads of 90-150 students per day, which makes it hard to find time for frequent graded writing assignments (Edyburn, 2020).
- Classroom writing practices are influenced by teachers' beliefs and knowledge (Graham, 2019).
- With no benchmark or objective standards in place for writing (unlike reading), it's extremely difficult to remove subjectivity (Edyburn, 2020).
- Some findings from specific research (Graham, 2019):
  - The primary audience for students' writing was the teacher
  - Writing involved little collaboration among students
  - Time spent preparing for high-stakes writing tests was excessive
  - Classroom resources for teaching writing were inadequate
  - Formative evaluation occurred infrequently
  - Motivation for writing was largely ignored
  - The needs of students with a disability or who were learning a second language were not sufficiently addressed.

Twenty years later, it seems the NCOW (2003) report still holds true that writing remains a neglected skill in our schools.

How does the research tell us about how writing instruction needs to change?

Below is a summary of the relevant findings related to changes that need to be made to improve writing instruction in our schools:

- More time and attention needs to be devoted to writing instruction, and reading and writing instruction needs to be balanced (Graham, 2019; Slavin et al, 2019). Timothy

Shanahan, an emeritus professor of education at the University of Illinois Chicago explains, "The body of research is substantial in both number of students and quality of studies. There's no question that reading and writing share a lot of real estate, they depend on a lot of the same skills" (Schwartz, 2023).

- Teachers need to have better training in implementing high quality writing instruction as research shows teachers devote more time and attention to teaching writing if they are better prepared to teach it, feel more confident in their capabilities to teach it, derive greater pleasure from teaching it, and consider it an important skill (Graham, 2019). Providing extensive professional development to teachers, in which they themselves experience the writing strategies they will employ produces good writing outcomes (Slavin et al, 2019).

- Goals for writing need to focus on using writing for real purposes and authentic audiences as well as writing in a more realistic fashion. Goals also need to address motivation, knowledge, and social contexts (Graham, 2019). Slavin also notes that motivation seems to be the key (Slavin et al, 2019). Students need to be able to adapt to contexts, meet challenges, and solve problems that are as yet unknown (Care & Vista, 2017).
    - A key finding of a study on the impact of an external audience on the quality of second graders' writing found that early elementary students produced significantly higher quality informative/explanatory writing when they were provided with an external audience (the local librarian) rather than when writing for the internal audience (the classroom teacher) (Block & Strachan, 2019).
    - Teaching and learning should be preparing students for the world outside of school where writers write for a variety of purposes and audiences. Schools need to provide students with opportunities to write and revise for specific purposes to external audiences (Block & Strachan, 2019).
    - We must provide students with the "why" for writing. While learning how to write is undoubtedly important, children (and teachers) must also experience the powerful reasons why we write: to foster social relationships, engage in civic responsibility, and share information. Meaningful writing instruction should support writing for an audience beyond the teacher and a purpose beyond a report card. Children should be given a variety of opportunities to express themselves and to understand how their writing can have real-world implications (Woodard, 2021).

- Education in the 21st Century is about skills—sets of processes (Care & Vista, 2017). An agreed upon set of skills, knowledge, and processes needs to be created for writing and there needs to be alignment across writing goals, curriculum, instructional methods, and assessment (Graham, 2019).

- Progression-based assessment tools need to be developed that provide performance information in language that describes that performance (Care & Vista, 2017) in asset-based language (Perks & Rosenfeld, 2022) that transparently communicates what writing looks like at every level. These progression-based writing tools will help to take the subjectivity out of writing assessment, will inform instruction and feedback, and will

help students to continue to improve their writing and measure progress and growth over time.

- Students benefit from step-by-step guides to writing in each specific genre (Slavin et al, 2019). The What Works Clearinghouse also highlights the importance of explicit writing instruction that varies by genre for both elementary and high school students (Barshay & Barshay, 2019).
- Students need explicit grammar and punctuation instruction but it should be taught in the context of their writing, not as a separate stand-alone lesson (Slavin et al, 2019).

The State of the Common Core State Standards

The Common Core State Standards (CCSS) were released in 2010. Initially, 46 states passed laws adopting the Common Core standards. There were four states that never adopted the Standards — Virginia, Texas, Alaska, and Nebraska. Many states have rescinded or repealed the CCSS legislation and have implemented or are currently in the process of developing new educational standards to replace Common Core. However, research conducted by Abt Associates for the *Massachusetts English Language Arts/Literacy and Mathematics Curriculum Frameworks Review* found that many states are not making substantial changes, and most of the changes to the ELA standards were clarifying changes or adding a concept or a skill (Norton et al, 2016)**,** amounting to more of a rebranding of the CCSS rather than a full replacement. The majority of writing assessments are either based on the three types of CCSS writing types—narrative, argumentative, and informational—or in response to reading. There is a growing body of research that indicates the types of writing teachers teach and how they teach writing is greatly influenced by the form of the state writing assessments (O'Neill et al., 2005).

Artificial Intelligence (AI)

It is impossible to write about the state of writing without addressing AI. There are so many questions looming about whether AI will replace writers and do all of our writing for us. No one can say for sure what impact AI will have on writers and writing, but it is highly probable, at the rate AI is advancing, this will be something to consider in 2030. If we view AI as another tool in a writer's toolkit, what are the possible ways to use AI to enhance our writing? It seems unlikely that AI will replace writers but, chances are, writers using AI will replace writers who are not using AI.

One possible way this might impact the 2030 administration is that there may be an additional AI tool that students can use like the spelling and grammar check that gives them feedback and suggested revisions based on the scoring tool, especially if there is automated scoring in use. It is highly likely that AI will be able to be used to score the writing assessments. Studies are underway comparing human ratings to automated essay scoring (AES) and cutting-edge AES tools can generate agreement levels similar or better than agreement levels between two humans (Kumar & Boulanger, 2021). Automation allows for simpler, more frequent assessments and can monitor student progress over time (Edyburn, 2020).

Recommendations for Updates to the NAEP Writing Assessment Framework for 2030

Introduction

Update purposes for writing and research to reflect latest data and findings and connect to the why for the NAEP Writing Assessment. Focus on equity and how you are making this assessment accessible and culturally relevant for all students.

The Role of Purpose and Audience

The research indicates that authentic purpose and audience are essential to writing instruction and improving writing outcomes for students. Keep the focus of the 2030 assessment on both of these areas and offer choices of possible audiences but also allow students to choose the audience.

Writing with Computers

All students should be given the opportunity to use a computer to take the test and should also be able to use a variety of commonly available tools, even in fourth grade and especially by 2030. We all use tools such as spelling and grammar checks to improve our writing. We should be designing assessments for the world we live in now and the type of writing we are doing now.

There should be a process to ensure students know how to use these tools prior to the assessment. Consider adding speech to text technology as a tool to help students overcome some of the barriers to producing a timed piece of writing in 30 minutes. This would also allow more time for editing and revising.

Time Per Task

The purpose of the NAEP Writing Assessment is to ascertain what students know and can do in a limited amount of time with limited resources. To this end, I disagree with the public comments suggesting that students need to be given additional materials to read and/or review in order to include evidence in their writing. While I agree that there are many standards and competencies that require students to incorporate evidence from credible sources into their writing, that is not the type of writing that is being assessed by the NAEP Writing Assessment. At every grade level, the CCSS asks students to, "Write routinely over extended time frames (time for research, reflection, and revision) and shorter time frames (a single sitting or a day or two) for a range of tasks, purposes, and audiences." The NAEP Writing Assessment is assessing students' ability to write in a shorter time frame (a single sitting) for a variety of purposes and audiences.

The addition of texts to read or graphs to interpret adds not only to the time it will take for students to complete the assessment, but also adds complexity to the task. Once you introduce the requirement of reading text into a timed writing assessment, you are no longer just

assessing a student's ability to write, but also to read and interpret information, which is already being assessed in the NAEP reading assessment.

Assessing Students with Special Needs

Separate students with learning differences from English Language Learners (ELL). Students with learning differences will have specific accommodations outlined in their Individualized Education Plans that can be provided as appropriate. There are a broad range of ELL students and I will defer to the recommendations in the public comments from TESOL International Association, WIDA, The Center for Applied Linguistics for multilingual learners.

Communicative Purposes for Writing

Based on the research, the three communicative purposes the NAEP assesses — to persuade, to explain, and to convey experience — are still the main three types of writing that are taught in schools and, therefore, should remain the three types of writing to be assessed.

Choice should be given in the topic students write about and there should be an opportunity to build background about the topics offered (e.g., watch a video about the topic). There needs to be a diverse group of people (e.g., from different cultures, races, locations) involved in designing the tasks to ensure they are topics everyone is familiar with and can write about. For example, the sample fourth grade task for *To Persuade* was about selecting a mascot. I have worked with many students, in all grade levels, who do not know what a mascot is and would have difficulty writing a persuasive essay about why one of those mascots should be chosen.

In addition to the importance of purpose and audience, student motivation is another key factor in writing quality. Although this assessment is important in understanding and evaluating the state of writing and writing instruction in schools across the nation, there is little motivation for students to complete this assessment. Consider how to increase student motivation and engagement.

How Responses Will Be Evaluated

It is important for students to know the criteria on which they are being assessed. This transparency allows students to try to achieve desired levels, helps them understand what the expectations are for a given task, and may increase their motivation and engagement in the task. We should be giving students every opportunity to succeed at the tasks on which we are assessing them.

A study examining the effects using instructional rubrics to help students learn to write found that 8th grade students who were given the instructional rubrics received, on average, higher scores than the students who did not receive the instructional rubrics. Additionally, students who received the instructional rubrics were able to identify more of the criteria by which their writing was evaluated (Goodrich Andrade, 2001).

Using the same rubric for all three grade levels makes the expectations at each grade level less transparent. If we give these rubrics to students, they will not be able to discern the expectations at each different grade level. Also, deficit language is not instructive and does not help students understand what the expectations are at the various levels. In addition, using the same six point rubric for all three grade levels also does not give us insight as to where the writing may be in relation to other grade levels and what growth and progress students can make in their writing over time.

Rethinking and reframing the scoring tool and sharing the scoring tool with students at the beginning of the task may improve the quality of writing and will also allow for a longitudinal view of student writing progress that may be more instructive than the 6 point rubric. We need to think of the tools we use for scoring as not just assessment tools but also as instructional tools for both students and teachers and the tools we design need to provide performance information in language that clearly and transparently describes that performance (Care & Vista, 2017).

In the revisions for the 2030 implementation, I recommend moving to a K-12 learning progression (example in Appendix A) for each area of writing that outlines what the writing expectations are at each level in asset-based and student-facing language. The language of the indicators at each level should increase in both complexity and rigor. Proficiency benchmarks for each grade level (4, 8, and 12) can be identified and labeled on the progressions. Students should be given time to review the scoring progressions in advance of starting the writing task. There should also be a way for students to self-assess on the scoring progression and submit those ratings with their writing.

Additionally, anchor papers will need to be rescored on the new progressions by a team of scorers to determine what writing looks like at each level and at the grade level benchmark levels. All scorers will need to be trained in using the progressions and assessed on their ability to score the anchor papers aligned to the progressions and the scores given by the benchmarking team.

Research indicates that there is a lack of common objectives for writing and the rating of writing is very subjective. NAEP could share these writing progressions nationally to help create a common set of expectations for improving student writing in these three areas.

Conclusion

The research that exists on writing instruction indicates that writing is still a neglected skill in our schools. More research needs to be done to learn how to best improve the writing of our students, but lack of time dedicated to teaching writing, lack of teacher training, and the lack of common expectations for writing are problems we know exist. The 2030 NAEP Writing Framework is an important way to continue to evaluate the quality of our students' writing. The recommended updates will help ensure the assessment is providing all students with the

opportunity to demonstrate their best writing abilities. Through the development of K-12 learning progressions that are shared nationally, the NAEP Writing Framework also has the potential to contribute essential tools to help improve the quality of instruction and writing in all of our schools.

References

Andrade, H. (2001). The Effects of Instructional Rubrics on Learning to Write. *Current Issues in Education*, 4. http://cie.asu.edu/ojs/index.php/cieatasu/article/view/1630

Barshay, J., & Barshay, J. (2020). Scientific evidence on how to teach writing is slim. The *Hechinger Report.* https://hechingerreport.org/scientific-evidence-on-how-to-teach-writing-is-slim/

Block, M. K., & Strachan, S. L. (2019). The Impact of External Audience on Second Graders' Writing Quality. *Reading Horizons*, 58(2), 68–93. https://eric.ed.gov/?id=EJ1230969

Care, E., & Vista, A. (2017). Education is Changing. It's Time Assessment Caught Up. *Stanford Social Innovation Review.* https://ssir.org/articles/entry/education_is_changingits_time_assessment_caught_up

Dillon, S. (2005). What corporate America can't build: a sentence. *The New York Times*, 1-4. https://www.nytimes.com/2004/12/07/business/what-corporate-america-cant-build-a-sentence.html

Edyburn, D. (2020). Poor Writing Skills Are a Crisis in Our Public Schools. *Texthelp*. https://www.texthelp.com/resources/blog/poor-writing-skills-are-a-crisis-in-our-public-sch/

Graham, S. (2019). Changing How Writing Is Taught. *Review of Research in Education*, 43(1), 277–303. https://doi.org/10.3102/0091732x18821125

Kumar, V., & Boulanger, D. (2021). Automated Essay Scoring and the Deep Learning Black Box: How Are Rubric Scores Determined? *International Journal of Artificial Intelligence in Education*, 31(3), 538–584. https://doi.org/10.1007/s40593-020-00211-5

NACE. (2022). Job Outlook 2023. Bethlehem, PA: National Association of Colleges and Employers.

National Center for Education Statistics. (2012). The Nation's Report Card: Writing 2011 (NCES 2012–470). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Commission on Writing. (2003). The neglected "R": The need for a writing revolution. Washington DC: *College Entrance Examination Board*. https://eric.ed.gov/?id=ED475856

Norton, J., Ballinger, S., & Ash, J. (2016). Massachusetts English Language Arts/Literacy and Mathematics Curriculum Frameworks Review, Final Report. *Abt Associates*. https://www.abtassociates.com/insights/publications/report/massachusetts-english-language-artsliteracy-and-mathematics-curriculum

O'Neill, P., Murphy, S., Huot, B., Williamson, M. (2005). What Teachers Say About Different Kinds of Mandated State Writing Tests. *Journal of Writing Assessment,* 2(2). Retrieved from https://escholarship.org/uc/item/0h5093qh

Perks, K., & Rosenfeld, R. A Strengths-Based Approach to Assessment Rubrics. (2022). *WestEd Bulletin*. https://www.wested.org/wested-bulletin/equity-in-focus/strengths-based-assessment-rubrics/

Slavin, R.E., Lake, C., Inns, A., Baye, A., Dachet, D., Haslam, J. (2019). Quantitative Synthesis of Research on Writing Approaches in Grades 2 to 12. Best Evidence Encyclopedia (Bee). *Center for Research and Reform in Education*. https://eric.ed.gov/?id=ED597167

Schwartz, S. (2023). 3 Takeaways About the Connection Between Reading and Writing Instruction. *Education Week*. https://www.edweek.org/teaching-learning/3-takeaways-about-the-connection-between-reading-and-writing-instruction/2023/02

Woodard, R. (2021). When Teaching Children How to Write, We Must Also Show Them Why to Write. *Chalkbeat Chicago*. https://chicago.chalkbeat.org/2021/7/29/22596821/cps-writing-instruction-skyline?_amp=true

Appendix A

Sample K-12 Learning Progression for the competency Express Ideas

ELA.2 EXPRESS IDEAS

I can clearly and effectively express my ideas (in written and oral form) for particular purposes and audiences, using diverse formats and settings to inform, persuade, and connect with others.

	LEVEL 2	LEVEL 4	LEVEL 6	LEVEL 8	LEVEL 10	LEVEL 12
ELA.2.1 **Identify a core message and audience**	I can choose the main thing I want to say to my audience.	I can choose **a central message for my product/performance**. **I can determine my audience and purpose.**	I can choose a central message (e.g., thesis, claim, story) for my product/performance. **I have specific ideas for tailoring my approach** for my audience and purpose.	**Drawing on diverse sources,** I can develop a central message for my product/performance **that connects to an important theme, idea, or issue in the world.** I have specific ideas for tailoring my approach for my audience and purpose.	Drawing on diverse sources, I can develop a central message for my product/performance that connects to an important theme, idea, or issue in the world. **I can articulate the impact I hope to have on my audience.** I have specific ideas **and strategies** for tailoring my approach for my audience and purpose, **and for achieving impact.**	Drawing on diverse sources **and original insight,** I can develop a central message that connects to an important theme, idea, or issue in the world **and that contributes to, or advances, the field or discipline.** **I can anticipate the impact these ideas will have on different audiences and/or my community, and craft my message in a responsible way.** I have specific ideas and strategies for tailoring my approach for my audience and purpose, and for achieving impact.
ELA.2.2 **Develop and organize the message**	I can choose the details I want to share. I can put the details in the order that makes sense. I can create an opening. I can create a closing.	**I can choose details and information that will help me achieve my purpose** (e.g., inform, persuade, entertain). I can organize my ideas **in a way that is easy for my audience to follow.** I can choose an opening **that gets my audience's attention.** I can create a closing that achieves my purpose **(e.g., inform, persuade, entertain).** **I can use words and examples that are relatable to my audience.**	I can choose **important details and/or evidence (e.g., claims, characters, plot, examples, reasons, rationale, actions) to develop my ideas** in support of my purpose (e.g., inform, persuade, entertain). I can organize my ideas **in a logical way.** I can choose a opening **that engages my audience (e.g., hook, setting a purpose, providing background, graphics or images, establishing a problem frame, using data)** I can create a closing that achieves my purpose **and highlights the importance of my topic (e.g. suggest a course of action, propose a solution, discuss a broader implication).** **I can apply tools, techniques and elements of craft** (e.g., word choice, phrasing, pictures, sound effects, tone, rhetorical devices) **to help engage my audience and connect with my message.**	I can choose **the most relevant and important** details, **descriptions,** and/or evidence to develop my ideas (e.g., claims, characters, plot, examples, reasons, rationale, actions) in support of my purpose, **addressing conflicting or alternative ideas or perspectives (when applicable).** I can organize my ideas **and supporting content around a logical structure/arc to achieve my purpose.** I can choose a opening that engages my audience (e.g., hook, setting a purpose, providing background, graphics or images, establishing a problem frame, using data) I can create a closing that achieves my purpose (e.g. resolution, conclusion, suggest a course of action, propose a solution, discuss a broader implication, invitation to engage further, raising questions for future inquiry). I can apply tools, techniques and elements of craft (e.g., word choice, phrasing, pictures, sound effects, tone, rhetorical devices) to help engage my audience **and effectively communicate my message.**	I can choose the most relevant and important details, descriptions, and/or evidence **to add depth or complexity to my ideas** in support of my purpose (addressing conflicting or alternative ideas or perspectives when applicable). I can use the **most effective structure** to organize my ideas and supporting content around a logical structure/arc to achieve my purpose. I can choose a opening that engage and **sets an intentional tone for my audience** (e.g., hook, setting a purpose, providing background, graphics or images, establishing a problem frame, using data). I can create a closing that achieves my purpose (e.g. resolution, conclusion, suggest a course of action, propose a solution, discuss a broader implication, invitation to engage further, raising questions for future inquiry). I can apply tools, techniques and elements of craft (e.g., word choice, phrasing, pictures, sound effects, tone, rhetorical devices) to help engage my audience and effectively communicate my message.	**I can explore a variety of diverse sources so I can choose** the most relevant and important details and/or evidence from to add depth or complexity to my ideas in support of my purpose (addressing conflicting or alternative ideas or perspectives when applicable). I can use the most effective structure to organize my ideas and supporting content around a logical structure/arc to achieve my purpose. **I can select from a variety of techniques and genres to create a compelling or provocative** opening and set an intentional tone for my audience. I can create a closing that achieves my purpose and **leaves a lasting impression or makes an impact on the audience** (e.g. resolution, conclusion, suggest a course of action, propose a solution, discuss a broader implication, invitation to engage further, raising questions for future inquiry). I can apply **sophisticated** tools, techniques and elements of author's/creator's craft (e.g., artistic pacing, complex reflection, engaging dialogue) that illustrate my creativity **and command of the genre,** and help amplify my message.

	LEVEL 2	LEVEL 4	LEVEL 6	LEVEL 8	LEVEL 10	LEVEL 12
ELA.2.3 **Prepare the medium**	With guidance, I can choose the best way to share my message with my audience (i.e., speaking, writing, showing).	I can choose the best **format** for reaching my audience (e.g., written story, presentation, video). **I can learn from high-quality examples and get ideas for my own product/performance.**	I can choose the **most effective format for my specific purpose and audience.** I can learn from high-quality examples and get ideas for **specific aspects of my product/performance.** **I can make choices about the features of my product/performance** (e.g., data, pictures, music, software tools) **and use of technology that help strengthen my message for my specific audience.**	I can choose the most effective format for my specific purpose and audience. **Using criteria,** I can learn from high-quality examples and get ideas for specific aspects of my product/performance. I can make choices about the features of my product/performance and use of technology that help me strengthen **or elaborate** my message and **positively impact my specific audience.**	I can choose the most effective format for my specific purpose and audience. **On my own, I can use criteria to source and** learn from high-quality examples and get ideas for my own product/performance. I can make choices about the features of my product/performance and use of technology that help me strengthen or elaborate my message and positively impact my specific audience. **I can incorporate design decisions that show evidence of my original thinking.**	I can choose the most **effective formats to tailor my products/performances for different audiences and/or purposes.** **Using criteria I've created,** I can source and learn from **exemplars, draw inspiration, and analyze format choice relative to purpose and audience.** I can make **differentiated** choices about the features of my products/performances and use of technology that help me strengthen or elaborate my message and positively impact my **different audience/s.** I can incorporate **novel** design decisions that **exemplify, or challenge, convention.**
ELA.2.4 **Finalize, practice and/or prepare**	I can use feedback to improve my work. I can make sure I've used complete sentences and punctuation in my speaking or writing.	**I can self-assess against criteria to identify areas for improvement.** I can use feedback to improve my product/performance. **I can edit my final product to ensure it meets the guidelines provided by my teacher.** **I can practice or rehearse my performance before I share (when applicable).**	I can self-assess against criteria to identify areas for improvement. I can use feedback to improve my product/performance **for my specific audience and purpose.** I can edit my final product to ensure **it follows conventions and standards for the chosen genre.** I can practice or rehearse my performance, **and make one or more adjustments to prepare for my performance** (when applicable).	I can self-assess against criteria to identify areas for improvement. **I can gather and selectively use feedback from others,** to improve my product/performance for my specific audience and purpose. I can edit my final product to ensure it follows conventions and standards for the chosen genre. I can practice or rehearse my performance, and make adjustments to prepare for my performance **and to ensure supporting materials or supplies are ready** (when applicable).	**I can solicit general as well as targeted feedback based on my self-assessment,** and selectively integrate feedback to improve my product/performance for my specific audience and purpose. I can edit my final product to ensure it follows conventions and standards **consistent with the professional world.** **I can participate in sufficient practice or rehearsal rounds to ensure a high quality performance,** make adjustments to prepare for my performance, and ensure **supporting technologies,** supplies, and materials are ready (when applicable).	I can solicit general as well as targeted feedback **from expert/s** based on my self-assessment, and selectively integrate feedback to improve my product/performance for my specific audience and purpose. I can edit my final product to ensure it follows conventions and standards consistent with the professional world, **or breaks from standard conventions for a specific purpose or effect.** I can participate in sufficient practice or rehearsal rounds **at the designated venue/location to ensure the highest quality performance,** making adjustments to prepare for my performance and **working with others to** ensure supporting technologies, supplies, materials, **and all other components of the performance** are ready (when applicable).
ELA.2.5 **Engage, respond and reflect**	When asked, I can answer questions about the ideas I shared. I can talk about how well I think my audience liked what I shared.	When asked, I can answer questions about my product/performance **and/or provide additional information to my audience.** **I can explain what I did well and what I would change the next time.**	I can respond to **a range of questions** (e.g., clarify, elaborate, critique) about my product/performance, **maintaining my composure.** **I can reflect on both my process and product to identify areas of strength and areas for improvement.**	I can respond to a range of questions about my work, maintaining my composure **and connection to the audience** (e.g., phrasing, tone, eye contact, references). I can reflect on both my process and product to identify areas of strength and areas for improvement. **I can reflect on the impact my work had on my audience.**	**With confidence and composure,** I can respond to a range of questions about my work, **choosing carefully selected words, examples, or resources to connect with my specific audience.** **I can evaluate** both my process and product to identify areas of strength and areas for improvement. **I can evaluate the impact my product had on my specific audience.**	With confidence, **integrity,** and composure, I can respond to a range of questions about my work, choosing carefully selected words, examples, or resources to connect with my specific audience. I can evaluate both my process and product to identify areas of strength and areas for improvement. I can evaluate the impact my product had on my specific audience, **including whether or not my final product impacted them in the way I intended, collecting and analyzing feedback from my audience when possible.**

Review of the 2011 NAEP Writing Framework
Submitted March 31, 2023
by Peggy O'Neill, PhD
Loyola University Maryland

Introduction

The Writing Framework for the 2011 NAEP reflected important shifts from the earlier framework. First, it emphasized a rhetorical approach to writing with each task identifying a purpose and audience. This situated writing as a communicative act. Second, it was a computer-based test for grades 8 and 12 with the expectation that by 2017, grade 4 would also be computer-based. For the computer-based administration, test material was provided via computer, and the students used word-processing software that included access to some common tools in writing their response.

All tasks, according to the framework, were one of three types of specified purposes: *to persuade, to explain,* or *to convey an experience, real or imagined*. These are common types of purposes for writing in academic, professional and civic contexts. They also align with the College and Career Readiness Anchor Standards for Writing in the *Common Core State Standards for Writing*. Audiences included a variety of options—depending on the task, purpose, and grade level—such as peers, general audiences, principal, and community leaders. The form of the writing was specified in some cases, and in other cases, was left open for students to select the form that they thought fit with the purpose and audience (this option was only available for some of the grade 8 and grade 12 prompts). These tasks are in keeping with writing as a communicative activity that is situated in specific situations while also acknowledging the wide variation in student knowledge and context.

Other changes recommended by the developers of the 2011 NAEP Writing Framework included using the same broad domains of writing in evaluating all responses across all purposes— *development of ideas, organization of ideas, and language facility and use of conventions*— instead of the mode-specific characteristics as was done previously. The evaluation of student responses in the broad domains would consider the purpose and audience of the task.

The 2011 NAEP Writing Framework also recommended the development of the *Profile of Student Writing,* based on a nationally representative sample of student responses at each grade level. The profile would use both qualitative and quantitative analysis of the samples, and it would be part of the report of the results.

Other aspects of the framework were not as different than those in earlier administrations. For example, each student responded to two prompts, with 30 minutes allowed for each. Students could not use outside sources, and the computer administration did not allow for internet access or other sources beyond the word-processing tools. Responses were evaluated using a holistic rubric.

Although the changes to the 2011 NAEP Writing Framework were significant, it did not embrace the full complexity of writing as a construct. For example, the 30 minute timeframe, without any outside sources or collaboration, does not represent the way writers write in most situations beyond standardized, impromptu writing tests. The 2011 NAEP Writing Framework developers, however, understood the limitations presented by the constraints of a standardized exam and acknowledged that the assessment would be measuring student writing in a limited, yet important capacity.[*] Keeping the time constraints and having each student respond to two different tasks allows for more valid and robust analysis of the results.

The *Profile of Student Writing* would provide more nuanced understanding of student performance, scoring, and results. The framework authors also acknowledged that the three specified purposes were not all inclusive; however, these are commonly used purposes in both academic, workplace, and civic situations. Limiting the test to three purposes allows students to write in response to different tasks, provides some variety in topic, purpose and audience, and provides standardization. This combination makes the analysis of the results more useful in terms of understanding strengths and weaknesses of student writing. The use of the computers allows for insight into how students compose using basic word processing software.

In short, the 2011 NAEP Writing Framework is a substantial move forward from earlier frameworks and is worth keeping for the next administration of the NAEP Writing Assessment. Now is not the time to do a revision or rewrite of the Writing Framework.

Primary Recommendation

After reviewing the 2011 NAEP Writing Framework, the public comments, the writing assessment reports on the implementation and use of the framework, and other related materials, I have concluded that the most important action is to conduct a NAEP Writing Assessment as soon as possible, *before* revising the current framework.

This recommendation is based on the following:

- The current framework is based in a rhetorical construct of writing that is aligned with how writing is defined by writing scholars and how it circulates in the world. This framework was a radical departure from the earlier NAEP writing frameworks; however, we have very limited information on how this shift impacted writing instruction and assessment. Given the issues with the results in 2017, we are behind in understanding the 2011 NAEP Writing Framework and its implications and limitations.

- Writing on a computer using word-processing software is (still) critical for writing in school and beyond in college, work, and civic contexts. Students from elementary through post-secondary need to be able to compose using word-processing software. We have limited information on the uses and accessibility of students writing with computers

---

[*] My understanding, from previous work with NAEP Writing Assessment and in reviewing the 2011 Framework, is that the time limits were non-negotiable although that may no longer be the case.

given the situation of the 2017 NAEP Writing Assessment. Getting a clearer understanding of this, especially post-pandemic, is essential for moving forward to revise the Writing Framework in an informed way.

- The recently released, and disruptive, artificial intelligence (AI) writing programs (such as ChatGPT) will impact our understanding of writing, what it means to be a writer, and the teaching of writing. The AI technology is still new and rapidly changing. It will be disruptive, no doubt. However, the consequences of the disruption is unknown at this point. Given the disparities that exist across schools, districts, and states, the disruption will be uneven. Making assumptions, at this point, is unwise because we are not good at predicting how technology innovations will play out in classrooms, workplaces, and other contexts. Waiting for the technology to "settle" and become more normalized will be necessary if we are to consider it in revising the writing framework.

- Given the changes in technology happening, the Governing Board needs to think more deeply about what we want to know about writing and how to find that out. Clarifying the purpose of the NAEP Writing Assessment will be necessary to revise the framework and implement it. Conducting a writing assessment using the current framework will help in clarifying the purpose and the methods needed. Determining the types of prompts, length of time, evaluation rubrics, and other aspects of the assessment should be based on what is discovered from a lengthy discernment, review, and research period. Waiting for that process to unfold is not tenable given the delays already incurred with the writing assessment.

- Since 2011, the last time the full results of a NAEP Writing Assessment were released, much has changed in the writing assessment landscape. The pandemic was the most disruptive in terms of assessments—and education more broadly—interrupting testing schedules and learning opportunities among other things. ESSA replaced NCLB, which gave states more flexibility in terms of assessment although it still requires statewide assessments in English language arts. The SAT eliminated the written essay option on its while the ACT maintained it. Across the country, there is a wide range of ways that ELA, and more specifically writing, is assessed, as reported by the Education Commission of the States. The changing assessment context highlights the importance of the information from a national sample of students that only the NAEP Writing Assessment can provide. In fact, some states, such as Georgia, consider NAEP as part of its summative assessment program.

Developing or substantially revising the current 2011 NAEP Writing Framework will be a labor- and time-intensive undertaking. Before doing that, we should administer one more round of writing assessments under the current framework. This will also allow for the impact of the AI technology to be better understood. It will allow us to better understand the word processing skills and accessibility of technologies to students moving forward. In revising the framework, we need to look back at what research and past results tell us as well as project into the future, given the long lead time between developing the framework and implementing it. We do not have enough information from the recent past especially with the the disruption of COVID to teaching and learning, to state and district assessments, and to educational research. We do not

have enough information about current technology access and use, not to mention the changing technologies such as AI.

In light of the need for robust information, I also recommend following the 2011 NAEP Writing Framework recommendation to create a *Profile of Student Writing*. This will be a useful resource for educators and policymakers across the nation as they consider the results and look to the future. The framework provides a compelling rationale and directions for the profile. This information will be helpful in future deliberations about the writing framework or test implementation.

While my main recommendation is to conduct the NAEP Writing Assessment as soon as possible using the 2011 Framework, I have addressed other issues and concerns identified in the information shared by the Governing Board including the February 9, 2023, memo from Holly Spurlock, Branch Chief for the National Assessment Operations, NCES to Sharyn Rosenberg, Assistant Director to for Assessment Development, NAGB; the NAEP Writing Assessment Framework Summary of Feedback Received During Initial Public Comment Period; and the full text of public comments gathered from November 29, 2022 through January 25, 2023.[†]

Secondary Issues for Future Framework Development

Although I am recommending the Governing Board administer the NAEP Writing Assessment before developing and implementing a new writing framework, I agree with some of the concerns and criticisms expressed in the public comments and in Spurlock's memo, which I address below. I do not, however, think that these issues are enough to delay the administration of the next NAEP Writing Assessment. In fact, results from the writing assessment would inform the new framework, and there could be useful studies conducted in conjunction with the writing assessment that would contribute to the future review of the writing framework. Results from school, teacher and student surveys also contribute important information for framework developers and implementation personnel. The survey data are also helpful for educational policymakers and practitioners.

Spurlock addressed issues concerning time per task and use of stimuli in implementing the current framework. Some of the public comments also addressed these issues. These topics have also been addressed by researchers in writing assessment.

Time on task is related to many other aspects of writing. In most situations, writers have the option to extend the writing over more than 30 minutes and to consult a variety of sources including peers/colleagues, research material, and other sources. Longer timeframes are needed for lengthier texts and for more fully employing an iterative writing process. The time constraint also limits the type of stimuli that is used to in the prompt as well as what students can be expected to do in terms of supporting their ideas with details, evidence, and analysis. These

---

[†] The public comments were not representative of the many different constituencies that have a stake in NAEP Writing Assessment. I think a delay in revising the framework would also allow for a more robust response from relevant professional organizations, scholars, teachers and community leaders.

issues are worth exploring in the context of other issues including administrative constraints, longitudinal trends, validity and generalizability.

Allowing 8th and 12th graders more writing time, as suggested could be useful but more study needs to be done of how that would impact what we can learn about student writing. The trade off—one sample in 60 minutes instead of two samples of 30 minutes each—may not be worth it because research shows that one sample is not enough for making generalizations about a student's writing ability since students perform differently on different genres:

> Because the ability to write differs from genre to genre, generalizable inferences are not appropriate. In order to draw conclusions about writing in general, writing assessment should rather include multiple tasks in multiple genres rated by multiple raters. (Bouwer et al., 2015 p. 98)

Other researchers have come to similar conclusions (e.g, Beck & Jeffery, 2007; Breland et al., 1987; Chen et al., 2007; Freedman, 1979; Graham et al., 2011; Graham et al., 2016; Huang 2008; Pringle & Freedman, 1985; Quellmalz et al., 1982 Schoonen, 2012; Van den Bergh, et al., 2012). While there are models that allow for more extended, authentic tasks in a large-scale assessment, these would not allow for the standardization that NAEP traditionally uses so that comparisons can be made across many other variables.

Another issue mentioned in the feedback addressed the prompt stimuli with some specifically suggesting more reading-based tasks for older students. Although many writing tasks in academic contexts include text-based reading, not all are based on extensive reading of a text. Combining more extensive reading as the stimuli for writing assessment will muddy the writing results, especially for students with lower scores. For example, if students struggle with the reading, how will that impact the writing? Did some students misconstrue the reading or have trouble with it? Or were some students strong readers but just slower, which translates into less time for writing? NAEP already does a reading assessment so including more extensive reading as part of writing is unnecessary. Writing, as Deborah Brandt (2015) argues, is important aspect of literacy that has been gaining significance and even surpassing reading in some ways:

> It is not unusual for many American adults to spend 50 percent or more of the workday with their hands on keyboards and their minds on audiences, spending so much time and energy in acts of writing, in fact, that their appetites for reading often wane. (p. 3)

The NAEP Writing Assessment needs to stay focused on writing and not get subsumed or confounded with assessments of reading.

In short, I am not convinced that allowing for more time and/or providing more extensive texts as the stimuli will be worth the tradeoff, but it requires more study. More study requires more time, and the Governing Board should move ahead with an assessment under the 2011 NAEP Writing Framework in the meantime.

Another important consideration that the Governing Board and future framework developers need to consider is the changing nature of writing in light of computer technological advances

and what the impact of that will mean for teaching and learning to write. Currently, NAEP assesses students' ability to generate original text within the specified constraints—including topic, purpose, audience, and time. The 2011 NAEP Writing Framework provides research-based guidance in doing that in a standardized format. It acknowledges that the construct of writing it assesses—on demand, general topic writing tasks—does not cover all of writing. What the Governing Board needs to discern is whether this partial view is enough, or if the assessment should try more accurately to mimic authentic writing situations—for example, allowing access to more technical tools, providing more extended time, or broadening the tasks (which would include purposes, audiences, forms and stimuli). These decisions need to be based on what research tells us about the writing construct, writing assessment, writing processes, and learning to write. Conducting that research will take time. This research needs to include results from another the NAEP Writing Assessment because too much has changed since the last test results were released. The issues with the 2017 writing assessment highlight the need for more information as soon as possible.

The NAEP Writing Assessment is an important part of the Nation's Report Card, but it has been missing for too long. State, district, and school educators, researchers, and policymakers need the information that only NAEP can provide. I urge you to focus on conducting the writing assessment as soon as feasible before revising the NAEP Writing Framework.

References

Beck, S., & Jeffery, J. (2007). Genres of high-stakes writing assessments and the construct of writing competence. *Assessing Writing*, *12*(1), 60–79.

Bouwer, R., Beguin, A., Sanders, T., & Van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing*, *32*(1), 83–100.

Brandt, D. (2015). *The rise of writing: Redefining mass literacy.* Cambridge University Press.

Breland, H. M., Camp, R., & Jones, R. J. (1987). *Assessing writing skill*. Research Monograph No. 11. College Board.

Chen, E., Niemi, D., Wang, J., Wang, H., & Mirocha, J. (2007). *Examining the generalizability of direct writing assessment tasks*. Report No. CSE. 718. National Center for Research on Evaluation, Standards, and Student Testing.

Freedman, A., & Pringle, I. (1980). *The writing abilities of a representative sample of grade 5, 8, and 12 students: The Carleton writing project (Part 2. Final report)*. Carleton Board of Education. https://eric.ed.gov/?id=ED217413

Freedman, S. W. (1979). How characteristics of student essays influence teachers' evaluation. *Journal of Educational Psychology*, *71*(3), 328–338.

Quellmalz, E. S., Capell, F. J., & Chou, C. (1982). Effects of discourse and response mode on the measurement of writing competence. *Journal of Educational Measurement*, *19*(4), 241–258. https://doi.org/10.1111/j.1745-3984.1982.tb00131.x

Schoonen, R. (2012). The validity and generalizability of writing scores: The effect of rater, task and language. In E. van Steendam, M. Tillema, G. Rijlaarsdam, & H. van den Bergh (Eds.), *Measuring writing: Recent insights into theory, methodology and practice* (pp. 1–22). Koninklijke Brill NV.

Van den Bergh, H., Maeyer, S., Van Weijen, D., & Tillema, M. (2012). Generalizability of text quality scores. In E. Van Steendam, M. Tillema, G. Rijlaarsdam, & H. van den Bergh (Eds.), *Measuring writing: Recent insights into theory, methodology and practice* (Vol. 27, pp. 23–32). Brill. https://doi.org/10.1163/9789004248489_003

IES Institute of
Education Sciences

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

To: Sharyn Rosenberg, Ph.D.
        Assistant Director for Assessment Development
        National Assessment Governing Board

From: Holly Spurlock, Ph.D.
        Branch Chief for National Assessment Operations
        National Center for Education Statistics

Date: February 9, 2023

Subject: Implementation Considerations for the Current Writing Framework

This memo describes implementation challenges NCES encountered in its work to implement the current NAEP Writing Framework.

The current writing framework describes a robust and valid model of large-scale, on-demand writing assessment. It defines writing as "a complex, multifaceted, and purposeful act of communication that is accomplished in a variety of environments, under various constraints of time, and with a variety of language resources and technological tools." (Framework, page 3) To appropriately assess this definition, the framework calls for the measurement of students' abilities to write for a range of audiences for three purposes: to convey experience, to explain, and to persuade. The emphasis on writing purposes and audience acknowledges the social and communicative goals of writing, is consistent with research on writing instruction, reflects most states' writing standards and outcomes, and aligns with objectives for writing at the college level (Framework, page 19).-The preliminary holistic rubrics give clear and concise guidance and have been the foundation for the rubrics and training sets used in the assessment.

The framework offers sufficient guidance to achieve these goals. However, it also introduces constraints that make some of the goals challenging to achieve. We discuss these challenges in implementing the NAEP writing framework below.

Framework constraints: Time per task and use of stimuli.

On page 3 of the Introduction, the framework describes writing as being "… accomplished in a variety of environments, under various constraints of time, and with a variety of language resources and technological tools." This suggests an assessment that asks students to write for different purposes, at different levels of complexity, using a variety of types of resources. However, in Chapter One of the framework, under the heading of Time Per Task, the framework states that the writing assessment "…will be administered as two 30-minute, computer-based writing tasks." This time frame limits item writers in two important ways: the kind of stimuli and topics we can ask students to write to and about, and the kinds of audience(s) we can ask them to address.

Time constraints place limits on the use of stimuli and on task topics. The 30-minute per task time constraint limits the variety, complexity, length, and number of resources (such as texts, video, images, etc.) that can be included in tasks. The framework requires that any stimuli be brief to avoid consuming too much writing time. This in turn limits the kinds of tasks students can be asked to complete. Since students do not have sufficient time to select and integrate information from sources to (for example) the explain or persuade, tasks must use topics that are likely to be part of students' daily experience or rely on common knowledge. Item writers must therefore work with a relatively limited menu of topics for explaining and persuading tasks, many likely familiar to students.

Time constraints place limits on addressing audience in tasks. The importance of writing to an audience is a core principle of the NAEP writing framework. The specifications accompanying the framework ask item writers to "Provide realistic persuasive scenarios that will enrich the writing situation and heighten the writer's awareness of audience." (Specifications, page 26). However, when it is not possible to use sources that can help to establish audience for students, finding "realistic scenarios" that will "heighten the writer's awareness of audience" is challenging. And without that clear sense of audience and helpful context, students' abilities to use their knowledge of their audience to shape their writing are limited. The more distant from their personal experience, the less likely a diverse population of students will be able to conceptualize audience. For example, the grade 12 task Big Discount asks students to write letters to their local council members arguing for or against the building of a big box store in their area. Very brief quotes from residents are used to set context. Being able to supplement this task with sources offering information about the issue and the varied perspectives of council members would allow item writers to better "enrich the writing situation and heighten the writer's awareness of audience.

Conclusion. For some kinds of tasks required by the framework, item writers are not unduly constrained by the 30-minute time frame. These are topics and audiences for which item writers know students can draw on their experiences and knowledge to write effectively. For example, grade 8 students can write very well to a task asking them to explain to adults what adults don't understand about young people their age. Grade 12 students can write very well in response to a task asking them to write to a potential employer conveying experiences that make them a good job candidate. To Explain and To Persuade tasks that focus on school-based or other very familiar contexts, especially for younger students, do not overly-constrain item writers. However, to allow item writers greater freedom in designing tasks that can better fulfill the framework's definition of writing, it would be helpful to allow block times to vary to allow for the use of more extensive sources and stimuli while still maintaining NAEP's 60-minute time limit.

NAEP Writing Assessment Framework

Summary of Feedback Received During Initial Public Comment Period

February 2023

Manhattan Strategy Group
4340 East-West Highway, Suite 1100
Bethesda, MD 20814
P: 301.828.1515
E: *info@manhattanstrategy.com*
W: manhattanstrategy.com

Table of Contents

Introduction

The National Assessment Governing Board (Governing Board) is responsible for developing and updating assessment frameworks for the National Assessment of Educational Progress (NAEP), also known as The Nation's Report Card. Frameworks guide the content of NAEP assessments.

In preparation for a potential future update, the Board conducted a preliminary review of the current NAEP Writing Assessment Framework between November 29, 2022 and January 25, 2023. Reviewers were asked to address three questions:

1. Does the NAEP Writing Assessment Framework need to be updated?
2. If the framework needs to be updated, why is a revision needed?
3. What should a revision to the framework include?

At the end of the reporting period, 21 individuals or organizations had submitted comments (see Table 1 at the end of this document for the full list of individuals and organizations who submitted comments). All comments submitted were reviewed and categorized by theme. The final list of themes included the following:

- General Positive Comments

- Content and Forms

- The Frameworks' Purpose and Goals

- Tools

- Technology

- Artificial Intelligence (AI)

- Accessibility, Accommodations, Multilingual Learners, and Equity

- Background Knowledge

- Time Allowed to Complete Assessment

- Scoring

- Formulaic Writing

- Test Administration

- Align to the 2026 NAEP Reading Assessment Framework

- Editorial Comments

- Update Research

- Other Comments/Questions

Overall, the respondents agreed that the current NAEP Writing Assessment Framework should be changed in some way; however, it was not always clear whether the suggested revisions imply a minor or major update. Although there were three submissions that indicated that administering the assessment as soon as possible was more important than revising the framework, each of these respondents also provided thoughts on future revisions for the framework, and those comments have been included in this summary.

Summary

The following summary describes each of the major themes included in the public comments on whether the NAEP Writing Assessment Framework needs to be changed, why revisions are necessary, and suggestions for revisions. Bulleted summary comments are included in each section.

General Positive Comments

There were 10 submissions that provided general positive feedback about the current NAEP Writing Assessment Framework. Summary comments follow.

- In reading the 2017 framework, I thought overall the information included in the framework was relevant to how writing is currently taught in schools.

- I like the chart that explains Criteria for Evaluating Responses (1.3). It also highlights that writing is to communicate rather than the focus on form and genre.

- Overall, I feel the NAEP Writing Assessment Framework is excellent and comprehensive. I served on the item review for reading and find the revisions to be excellent. I am certain the writing revisions will also be valuable and excellent.

- I appreciate the thoughtful consideration of the types of writing to be assessed as well as the effort to provide students with writing topics that are relevant to the tested age groups.

- Appreciate the contextual variables.

- In reviewing the framework, I found it to be quite well done and appreciated the thorough consideration of the majority of aspects of writing and assessment.

- I found the overall framework thoughtful and well outlined. I appreciated the considerations for what is age appropriate since I work with elementary-level assessments. I found the scoring guides to be descriptive and inclusive.

- I feel the framework is thorough and taps into the knowledge and skills in ways that will provide the type of information that will help researchers in the field of writing understand what direction new research needs to go in, especially as technology advances in the future.

- We commend NAEP for conducting a study to determine the practicality of computer-based writing assessment for Grade 4 examinees. We encourage NAEP to continue to study the computer-based writing properties of responses from these younger examinees.

- I believe that the NAEP Writing Assessment Framework largely captures the modes of writing (to explain, to persuade, to convey real/imagined experience) that are most reflective of the fundamental skills writing students will need. I am mostly suggesting that innovations are needed in the assessment's constructs.

Content and Forms

Ten out of 21 submissions included comments on Content and Forms. Several of the comments encouraged revisions to the Writing Assessment Framework by making the assessment more relevant to students by incorporating contemporary writing, such as blogs, social media, texting,

etc., although there were cautionary warnings to not overemphasize informal writing. Allowing students to use blended writing as a response form was also recommended, along with asking students to write more about their personal experiences to encourage more authentic writing. A comment recommended reconsidering the audience that students are directed to write to on the assessment. Currently, fourth graders are asked to write to peers, but that is a skill that also benefits students in Grades 8 and 12. Another comment encouraged the framework writers to add questions that address the writing process (either on the assessment or through the contextual variables). Summary comments follow.

- Purposes of writing should remain the same, but consideration for the examples/forms students will experience on the NAEP should be considered to align with more contemporary writing, such as blogs and online articles. Students should also have a choice of topics so they can select topics relevant to their reality.

- Consider expanding the mediums through which students can write. For example, one of the current writing tasks is "persuade a classmate to read your favorite book." Students can successfully achieve this through multiliteracies such as videos, tweets, and TikTok videos.

- Because students use various media (e.g., texting, email, social media) to communicate, NAEP needs to take that into consideration. However, NAEP should not become an assessment of informal writing. The reason students are assessed in Grades 4, 8, and 12 is to determine their preparedness for what comes next in school, college, and the "real world," which includes readiness to communicate in writing "properly," in addition to other modes of communication.

- Current writing instruction is moving away from the distinct task types. Students often are called upon to share a personal experience (narrative) as part of their evidence to support their stance when writing an argument (persuade). Consider tasks in which students are given opportunities to demonstrate their skills and progress across multiple task types.

- Most state standards require students to use evidence to inform their writing; therefore, the framework should explicitly require students at each grade level to use evidence in their assessment responses. Prompts should include articles, images, and graphics to provide data and information that each student can incorporate in their assessment responses.

- Blended writing is not emphasized, although most writing includes some level of blending forms. Exhibit 1.2 shows that the forms matter little since the audiences and examples reflect blended audiences and tasks, and the broad domains of assessment shown in Exhibit 4.1 support the idea of blended writing being a viable genre and approach. It is not clear if blended writing is considered during the evaluation of the writing task.

- Reconsider or redefine the construct of writing from a "quill and ink model" to a 21st-century model that includes digital citizenship and multimodal forms of expression. Modern communication is increasingly supported by visual and digital content, lower density of text, and nonlinear formats with attention to digital citizenship and multimodal literacy. Consider revisions to the formats and tools utilized in the framework and writing assessment to match the demands of modern communication in a digital and globally interconnected context.

- There is space for the NAEP Writing Assessment Framework to expand to encompass more real-world writing situations and to provide students with the tools that they have mastery over to demonstrate their abilities to write and communicate in this era.

- Students need to be aware that they may use different approaches during the test.

- Broaden the audience that students write to at each grade level; specifically, allow all grade levels to write to peers. Only Grade 4 students are asked to write to peers. Students in eighth and 12th grade are requested to write to more authoritative audiences. Measuring students' abilities to effectively communicate to peers across all age groups would enable students to participate more fully in their social worlds.

- A revised framework should emphasize the importance of the writing process to understand student writing, and the student questionnaire should include questions about the writing process. The framework should acknowledge that while the final product is what readers have access to when reading a piece of writing, there are multiple stages of thinking, drafting, rethinking, revising, and editing that culminate in the final piece. Knowing more about how students are learning to write can help educators, school leaders, and policymakers continue or improve practices focused on the writing process.

The Framework's Purpose and Goals

From the 21 submissions received, nine included comments on the framework's purpose and goals. Suggestions for revisions included making the overarching goal clear (i.e., communicate effectively in writing across a variety of topics and formats). Several comments discussed ensuring that the assessment covered a variety of topics and formats. For example, respondents recommended adding "technical writing" and "writing to solve problems" to the assessment framework. Comments also focused on ensuring that the assessment provided students enough opportunities to demonstrate their knowledge by asking students to respond to more but shorter writing tasks. Additionally, students should be allowed the opportunity to provide evidence that they have the necessary skills to write using narration, description, classification, and evaluation. A couple of comments mentioned providing writing tasks that align to college and career readiness. Although the comments indicated the importance of preparing students for postsecondary writing expectations, two concerns were raised. First, for fourth and eighth graders, the focus should be on preparing for middle and high school, respectively. Second, an argument also was made that writing tasks should align with the types of writing students experience in everyday life. Summary comments follow.

- The framework mentions three goals (moving beyond formulaic approaches, using word processing software, and completing a writing task on-demand). A key goal is for students to communicate effectively in writing across a variety of topics and formats. The overarching goal should be stated explicitly.

- The revision team should consider that the statement about writing to communicate is more important than the narrow definition and mindset of persuade, explain, and convey real or imagined experience.

- A revised framework should be more closely attuned to the writing in the disciplines in Grades 8 and 12, including clarifying the difference between general and discipline-specific writing. In the current framework, NAEP is sending conflicting signals to the public about the purpose of the assessment and the kind of writing that should be taught in different disciplines. Including interdisciplinary topics to write about but not defining interdisciplinary expectations muddies already opaque instructional territory.

- There should be at least one writing task, preferably more, for each of the three communicative purposes.

- A primary purpose is to prepare students for postsecondary expectations. While this should be emphasized for the 12th-grade students, fourth graders need to demonstrate writing skills to benefit them in middle school and eighth-grade students need to demonstrate writing skills that will serve them in high school.

- The explanations for the descriptors need to be updated. This could be resolved by including more examples for each purpose. For example, "to persuade" does not always have to mean to change someone's viewpoint. It can also be to develop an argument, to analyze ideas, or to present a position. Tell readers that these purposes are not always mutually exclusive. More description could be added to the table on page 11 that compares the 2011 and 2017 frameworks.

- Current research on large-scale assessments shows too few separate writing items to have a wide range of task difficulty. A revision should include having a greater number of shorter writing tasks, possibly focusing on paragraphing at Grade 4, rather than having students "write a letter" or "write an essay."

- For students to be more well-rounded and suitably prepared for this ever-changing landscape, the need for technical writing becomes more crucial. Where does conveying specialized information fit into this framework? Students need the skills to do more than just report information coherently; they need to be able to guide and instruct the reader in application. This is a skill the majority of students will need to be college and career ready.

- It is important to make the writing authentic and aligned to the styles that are "college and career" ready, but the writing must also match the style of writing students experience in everyday life.

- Anchor the Writing Assessment Framework in universal design for learning (UDL) and its three key principles.

- Specifically assess writing ability using narration, description, classification, and evaluation skills as related to the purposes for written communication: persuasive, experiential, expository, and writing in the arts.

- Add "writing to solve problems" as a purpose. It is reasonable to predict that writing will be an avenue to solve complex and integrated civic, social, scientific, and economic problems. This change would also align with the 2026 Reading Assessment Framework, with the Reading to Solve a Problem (RSP) blocks.

- Limiting discussion around reasons to write for economic purposes may limit students' motivation to write, or to appreciate learning to write. Undeniably, there is a relationship between education and economic success. However, there are other kinds of success about which both society and individuals' care. A rationale that includes measures of success beyond the financial may be more appealing and convincing. Emphasizing that there are myriad ways in which writing may contribute to a variety of endeavors may encourage student understanding of why writing instruction and assessment receives and deserves so much attention.

Tools

Six submissions included comments about students' access to tools on the Writing Assessment; however, the respondents did not agree about the use of tools. Two comments specifically mentioned that tutorials about using the tools were not included in the framework. Two other comments argued for the addition of more tools for students, such as ClipArt, dictionaries, and digital representations. One comment warned that students may not use the tools if they are not aware of them. Final comments raised concerns about including tools in the assessment because the inclusion of tools may indicate to students that these skills are not necessary to learn. Summary comments follow.

- Include the implementation/embedding of tutorials that encourage the use of and how students use word processing and spelling/grammar tools provided.

- The 2017 NAEP Writing Assessment Framework includes the use of online tools, such as spellcheck and other typical word processing tools. It does not include a description of any tutorial about the use of those tools embedded within the NAEP.

- One way that the framework may open more creative approaches to writing as a form of communication is by enabling composing tools, such as ClipArt, font color, and the Internet as a database. While the current framework suggests that these composing tools are "irrelevant to an assessment of writing or distracting to students" (page 30), these tools are the means through which people in society express themselves. To deny students access to this during the assessment does not result in an accurate representation of what students can achieve.

- All students will benefit from the option to access the assessment with scaffolding, using tools and supports, such as: (1) accessing a reference, such as a dictionary or thesaurus (English or multilingual), and (2) using multimodal means of expression, including images and digital representations.

- Often students overlook or do not use the tools available to them unless they are specifically directed to make use of them.

- The availability of composition tools and research included in the framework as noted in Chapter 3 acknowledges improvement in the development of ideas due to use of those tools, which seems to conflict with the measure of a student's skills. There is a similar concern with the Language Facility and Conventions Domain.

- In Chapter 3, remove editing, spelling, and grammar tools from the design because it indicates that students do not need to learn these skills. If the tools remain in the design, then the text should be revised to reflect that expectation is a revised draft.

- Tools, such as spellcheck, have led to the belief that the machine will "fix" any grammatical problems in the document. NAEP needs to ascertain whether student writers will edit their own writing without the benefit of these types of tools.

Technology

Four submissions included comments about the NAEP Writing Assessment Framework and its relationship to technology. The first two comments focus on updating the framework to account for the technological advances made over the past couple of decades. When the framework was

written, students did not have access to the technology that schools presumably provide now. However, one comment cautioned that some students still may not have access to as many technological resources as their peers. Other comments suggested that the updated framework include innovative technology, and another comment asked for clarification as to whether students would use a word processor or an NAEP interface. Summary comments follow.

- The Writing Assessment Framework looks at students' ability to use the computer. Most schools use computers regularly and many offer one-to-one computers.

- The framework needs to be on technological innovation and how the changes affect writers in the workplace and in educational institutions. Children and youth need to be competent in using, understanding, and applying the current, innovative technology. In 2000, we needed to teach kids how to use computers and digital resources. In 2023, it means teaching kids about revolutionary AI technology.

- A revised framework should expand technology use beyond word processing. Instead of using technology as a proxy for writing quality and restricting the use to word processing, the framework should consider how technology could be used to improve writing content quality, researching, digital collaboration, fact checking, and multimodality. NAEP may want to consider allowing students to integrate multimodal aspects (e.g., video, audio) to their writing. Such additions are a more accurate reflection of the kind of writing writers do in the 21st century, would allow for technology to develop an idea, and could be evaluated.

- The framework should include information about whether examinees will have the option of word processing tools to use, or if they will have to enter their responses in an interface designed by NAEP. There are advantages and disadvantages to either approach. An NAEP-designed interface would provide a consistent, standardized set of tools for examinees. However, examinees may feel more comfortable using a word processor with which they are familiar.

- Despite the growing access that students across grade levels have to technology, disparities remain both in access and in the consequent comfort with and fluency in using technology, especially in elementary school. Students may use computers frequently to type text, but that does not guarantee their familiarity with planning, editing, and formatting tools that they are expected to use in the assessment. Consider providing fourth graders with the option to write their piece by hand instead of typing on the computer.

Artificial Intelligence (AI)

Three submissions addressed the advent of AI and the NAEP Writing Assessment Framework. First, a comment addressed that students need to know, understand, and use AI ethically. Second, although one comment indicated that the current framework did not need updating, the commentator raised several questions regarding AI that will need to be considered for NAEP in the future. A third comment raised security concerns for NAEP as AI becomes more prominent and accessible. Summary comments follow.

- The Writing Assessment Framework needs to teach students to know, understand, and use AI ethically. Education continues to call for technological innovation in the form of AI (e.g., ChatGPT). Educators need to understand and teach this technology to children.

- Although the framework does not need to be updated, there are questions concerning AI that educators need to consider:

- With the advent of AI (such as ChatGPT), how can we prevent students from attempting to have AI write essays for them?

- Will students learn how to use AI to compose essays for them that will achieve their communicative purpose effectively (e.g., similar to how they use calculators to solve math problems)?

- Will AI's ability to write essays change what we want students to know and be able to do in the production of written text?

- With the advent of AI "bots" that can write, test security may have to deal with new challenges. NAEP will need to be immune to the many forms that cheating can take. Students may be required to check devices, such as phones and watches, at the door.

Accessibility, Accommodations, Multilingual Learners, and Equity

Eight submissions included comments regarding revisions to accessibility, accommodations, multilingual learners, and equity. First, respondents asked for consideration in providing the assessment in multiple modalities and to provide accommodations, such as reading the prompt aloud. Second, additional concerns were raised about Internet access, especially for students in rural areas. Third, a concern was raised about Grade 4 students' ability to use the computer to write. If they are unfamiliar with using a computer to write, their writing ability may be inaccurately assessed. Two comments addressed multilingual learners and students with disabilities. For example, recommendations were made to acknowledge crosslinguistic transfer, to factor in how culture influences language on the Writing Assessment, and to use translanguaging. To build on the theme of equity, comments also suggested allowing students to address universal themes and have choices in responses to allow multilingual students more opportunities to showcase their writing skills. Comments also suggested multiple bias and sensitivity reviews. A recommendation was made to look at multilingual learners, multilingual learners with disabilities, and students with disabilities as different subgroups; accommodations should only be used for students with Individualized Education Programs (IEPs). A more general comment about ensuring a more equitable experience for all students suggested that students be provided with an option to access knowledge prior to beginning a writing task. Summary comments follow.

- The NAEP Writing Assessment should be administered by computer, but the framework should establish processes to ensure accessibility and equity in how the assessment is administered to all students. This may include providing the assessment in multiple modalities to accommodate the needs of under-resourced schools and/or for students with learning differences.

- Accommodations should be updated to emphasize computer-based instead of paper-based accommodations (i.e., large-print booklets, Braille versions of the assessment).

- An accommodation allowing the writing prompt to be read aloud should be added.

- There is concern about the availability of Internet access in rural areas of the country, which may indicate that students with limited access may have a more challenging time using the computer to write. More research may be needed to determine if access to the Internet may impact student scores.

- Using computers may confound the ability to assess students' writing ability. Not all fourth graders use computers equally well. There are access issues with low-income students who may not have adequate keyboarding and computer knowledge and skills.

- Incorporate opportunities for multilingual learners to access and utilize their rich linguistic and cultural resources during the assessment. For example, crosslinguistic transfer (use of two languages) should be acknowledged in writing for specific purposes (i.e., for fluency and depth of expression). Acknowledgment of how this should be treated in the framework is needed. Linguistic varieties should be included in both prompts and scoring.

- Culture influences language and should be factored in this area of assessment.

- Encourage and provide inclusive opportunities for multilingual learners and other minoritized students to present perspectives that are representative of their multicultural orientations. This can be achieved in multiple ways: (1) topics should draw from universal themes that have wide-ranging applicability across cultures and student experiences, (2) provide choices in prompts to address multiple cultural orientations, and (3) have multiple rounds of bias and sensitivity reviews with multiple stakeholders from diverse communities.

- Secure resources to review assessments written in multiple languages. As multilingual learners are learning in multiple languages and bring diverse linguistic resources to their writing, the NAEP Writing Assessment Framework should match the students' mode of instruction and enable multiple means of expression that include students' rich linguistic resources and language variations. The NAEP Writing Assessment Framework should: (1) approach writing assessment from a multilingual perspective and (2) have comparable forms in multiple languages.

- Treat multilingual learners and students with disabilities independently as they represent distinctly different groups of students. Multilingual learners, multilingual learners with disabilities, and students with disabilities are three distinct student populations, each with unique resources. Embarking on the redesign of the framework should center on prioritizing characteristics of these groups that can potentially positively impact the results rather than retrofitting the existing framework through accommodations. While accessibility through Universal Design for Learning (UDL) is to be afforded to all students, the use of accommodations is only a legitimate and valid route for students with Individualized Education Programs (IEPs).

- Disaggregate and report data using Every Student Succeeds Act reporting categories and show growth of multilingual learners of English (MLEs) over time in cohorts.

- Allow translanguaging. A revised framework should acknowledge that translanguaging-writers use their entire linguistic repertoire, including the range of languages and dialects they speak. The scoring rubrics should be updated to reflect the accepted use of translanguaging. Translanguaging is an asset to student writing, which will also expand equitable opportunities for student writers to show their strengths. Considering there is only 30 minutes where students produce on-demand drafts, multilingual learners should be encouraged to use their natural translanguaging writing process, languages, and dialects during planning and drafting.

- NAEP could provide a more equitable writing experience by providing the option for all students to access knowledge for the task before writing.

Background Knowledge

Four of the 21 submissions included recommendations to take students' background knowledge into account to ensure equity on the assessment for all students. Most of the comments raised concerns that students who did not have the appropriate background knowledge (e.g., understanding of how to write to a state legislature) would not be able to adequately complete the writing task. This could be addressed through providing more text-based responses. For students who are assigned a form, there was a concern that the student may not be familiar with the assigned form. A final comment indicated that the questions did not seem representative of different cultures or interests. Summary comments follow.

- A student's ability to read and comprehend a text is dependent upon a number of factors, including a student's decoding ability, ability to read fluently, background knowledge, sociocultural background, and motivation/engagement with each task. When all these variables are not taken into account, their ability to write based on their comprehension of the text may not be assessed accurately.

- Students may not have enough background knowledge to develop their points to the extent or depth required by the tasks. Short texts to build background knowledge could be provided.

- Students, specifically in Grades 8 and older, may not have the background knowledge to write out their understanding. For example, one task is "Take a position and write a response persuading members of your state legislature to support your position on whether or not protected land in your state should be opened to energy companies for drilling." These decontextualized topics make writing much harder. The Writing Assessment Framework is missing understandable contexts for writing. Contexts should be accessible from readings or video clips.

- A revised framework should provide opportunities for students to build knowledge on the topic before they begin writing. While the 2017 framework allows students to include information from their own reading, observations, and experiences and respond to short reading passages or visual stimuli, the sample tasks do not provide evidence that students have enough access to information to write anything meaningful about the tasks.

- A caution was raised for students who are directed to use one specific form. Students who are instructed to write an editorial, for example, will not be able to use their writing skills unless they know what an editorial is and have had experience reading and writing them. Audience is something else that students need background knowledge of (e.g., state legislature). Would they access content for writing to the state legislature from their civics knowledge, or would they become distracted by what they hear from social media?

- Specific scenarios are sometimes distracting for student writers, particularly those with lower ability. This is not to say that the research cited in the framework is irrelevant or that the ability of students to adjust language to specific audiences is not a valuable skill, but that specificity may distract some students from demonstrating their true ability.

- Given the 30-minute time limit for the assessment and the diversity of background knowledge among students even at the same grade level, it is likely that the diversity in approaches to writing remains an aspirational goal but not a requirement for demonstrating proficiency in this assessment.

- More text-based responses need to be included. This means that students do not have to have background information regarding the prompt material and instead can rely on the provided text for details and examples, which also makes for a more equitable writing experience. (The students may need more time to complete the assessment, and the texts may need to be read aloud to be equitable and to better assess writing versus reading ability.)

- The tasks seem geared toward White, middle-class students (i.e., fourth grade is focused on mascots, eighth grade is focused on achieving goals, and 12th grade is focused on community). These topics do not seem to have been selected as representative of different subcultures or interests. Consider giving writers a two-choice prompt.

Time Allowed to Complete the Assessment

Seven submissions regarding the timing of the assessment were received. Overall, comments recommended increasing the amount of time for students to complete the NAEP Writing Assessment, especially for students in the older grades. One comment, however, indicated that 30 minutes seemed reasonable, while another comment suggested providing a rationale for the 30-minute time limit. Summary comments follow.

- Currently, the framework states that students will be given two 30-minute writing tasks. After reviewing the assessment for Grade 12, a suggestion is to allow at least 45 minutes for each task. With a more detailed prompt, students need time to process the information, create a plan for writing, and execute a draft. More time on the assessment would give students the opportunity to demonstrate critical-thinking skills as well as writing expertise.

- A revised framework should reflect more reasonable expectations for writing that account for all parts of the task: understand the task, purpose, and audience; determine the best format; draft; revise; edit; and publish. The complexity of the task and the time it takes to achieve it is not accounted for as the task becomes more challenging in Grades 8 and 12. Not allowing appropriate time for students to think limits students' ability to truly show their capabilities. NAEP should consider extending the time for all students, especially in Grades 8 and 12.

- In general, the assessment tasks do not seem tightly aligned to the note about the "pace of written communication" in the introduction. Encouraging the application of a speedy approach to writing may not be desirable. If a goal of education is to develop critical thinkers and lifelong learners, pointing to a habit of quickness to explain the importance of writing is illogical.

- Thirty minutes is a short amount of time for students to decide how to approach the writing, select some key ideas and their details, and compose them as a writing piece. Rather than lengthening the time, which would challenge implementation integrity and student fatigue, consider collecting student writing produced in actual 30-minute sessions and analyzing it for depth against the rubrics mentioned in the framework.

- Research or a rationale for the 30-minute allotment for each task should be included.

- Thirty minutes seem to balance testing time with the need for time to pre-write and edit.

Scoring

Seven of the 21 submissions related to scoring. Suggestions included disaggregating the rubrics by grade level and transitioning to an analytic rubric rather than a holistic rubric. Comments indicated that it is unclear how writing is evaluated with the holistic rubric, and several suggestions were made on how to update the rubric. Another suggestion was made to consider automated scoring. Finally, two questions were submitted indicating a need to clarify scoring in the NAEP Writing Assessment Framework. Summary comments follow.

- It would be beneficial if the rubrics were disaggregated by grade level, with age-appropriate criteria for each grade. Separating rubrics by grade level would make explicit some important developmental considerations. For example, a fourth grader may not have the cognitive capacity to add creative or figurative language, but it may be an expectation for a student in Grade 12.

- A revised framework should consider an analytic instead of a holistic rubric. At a minimum, more clarity is needed around how the holistic rubric evaluates the expectations in each category. An analytic rubric may provide insight into the areas of strength and growth.

- With the holistic rubric, it is unclear how writing is evaluated and trends are identified.

- A thorough review of the evaluation of responses is recommended, including the rubric design process, the rubrics, reader demographics, reader selection, reading training methods, and training materials.

- The training process described indicates that anchor papers will be specific to each grade level, and that those papers will flesh out the meaning of the rubric. Clarification is needed about whether the words in each rubric need to change or whether the anchor papers will define rubric terminology.

- It would be helpful to have student samples to illustrate what writing might look like. It would also make it easier to conceptualize a holistic score.

- Given that one of the stated goals of the framework is "to assess students' writing using word processing software with commonly available tools," a revised framework should precisely articulate what is being measured through or with technology use. As it is currently written in the framework rubric, it is unclear how technology use is being measured. A revised framework should include clear and precise explanations of how technology use is measured.

- The 2017 NAEP Writing Assessment Framework's focus on approaches to thinking and writing in relation to purpose and audience acknowledges a variety of ways to demonstrate writing ability, but this is not reflected in the preliminary rubrics. The format of the rubrics encourages an analytic process and contradicts the holistic scoring aims, resulting in ambiguity about how to use the rubric and measure the construct.

- Given that responses will be collected via computer-based assessments, NAEP may want to consider automated scoring. While automated scoring may not be appropriate for this upcoming assessment, there would be sufficient time for automated scoring professionals to

review the typewritten responses and hand scores from this administration and determine if automated scoring models are appropriate for future administrations.

- Does an achievement level of *Basic* include students who do not demonstrate partial mastery? Is there such a thing as not scorable (i.e., students who write little or nothing)? Does the holistic score of 1 demonstrate partial mastery? How will this be interpreted?

- If the goal of writing for all learners is to communicate, will the fourth-grade students assigned a "form" be penalized if they select a form that is not assigned?

Formulaic Writing

Two submissions related to formulaic writing, but they differed vastly. One respondent liked the idea of moving beyond formulaic approaches, while another respondent raised concerns about moving away from formulaic writing prompts. The comments also indicated that understanding how teachers approach formulaic writing would help define what "moving beyond formulaic approaches" means. Another comment recommended that scorers would need examples of how students may approach a non-formulaic writing task. Summary comments follow.

- I appreciate the goal to encourage student writers to move beyond prescriptive or formulaic approaches in their writing.

- The goal of encouraging "student writers to move beyond prescriptive or formulaic approaches in their writing" is problematic because formulaic approaches help many students learn to write successfully. This goal should be eliminated or include a discussion about how formulaic approaches are often an excellent way to teach several different types of writing genres.

- For non-formulaic approaches (which is recommended for 12th grade only), the training materials for scorers need to show multiple examples of the possible approaches that a writer may bring to the task.

- The need to move beyond formulaic approaches in writing is mentioned several times in the framework, yet few descriptions are given to describe what that idea means. Investigating what teachers consider to be formulaic writing and how it compares to what they are actively teaching to their students during writing would be enlightening.

Test Administration

From the 21 submissions, three focused on test administration. Two of those comments recommended administering the NAEP Writing Assessment as soon as possible without updating the framework. The final comment confirmed that fourth graders should be able to complete the assessment using a keyboard. Summary comments follow.

- Available resources should be devoted to deploying the Writing Assessment Framework sooner than 2030. No data have been collected since 2011; if the test is not administered until 2030, that is a gap of almost 20 years. This also impacts trend.

- The NAEP Writing Assessment should be given as soon as possible. The Governing Board should commit to administering the existing Writing Assessment to provide data to states and districts, ensure the public has information about student performance, and clearly convey the centrality of learning to write. It is more critical to ensure that there is clear data

on students' writing ability, particularly in the face of current technology that can write like and for humans.

- It is reasonable to have Grade 4 students use keyboarding for the Writing Assessment.

Align to the 2026 NAEP Reading Assessment Framework

The NAEP Reading Assessment Framework was recently updated, and three submissions suggested that the NAEP Writing Assessment Framework align to the changes made to the Reading Assessment Framework. Summary comments follow.

- Connect to the 2026 NAEP Reading Assessment Framework. The 2026 Reading Assessment Framework requires students to produce writing to demonstrate reading understanding and expects them to read discipline-specific texts, use technology, and solve problems. Connecting the Writing Assessment to the revised Reading Assessment may allow for a better, more seamless experience for students and more accurate information about student reading and writing ability.

- The 2026 Reading Assessment Framework assesses reading outside of traditional structures. The next iteration of the Writing Assessment Framework should align with the Reading Assessment Framework and focus on assessing student writing in innovative ways.

- Students should be expected to supplement the thoughts they bring to the assessment experience with information they read during the assessment on tasks beyond writing to convey experience. Though reading and writing are separate assessments, skillful writing often requires reading. This would align with the proposed changes to the 2026 NAEP Reading Framework, specifically the new "Use and Apply" comprehension target, which asks "readers to use information they acquire through reading to solve a problem or create a new text." For example, after a set of commentaries, readers might be asked to produce a blog-type measure for a public audience that captures the most relevant information or offers argument about an issue.

- The Reading Assessment Framework includes the use of multimodal texts and an expanded understanding of literacy, which should be included in the Writing Assessment Framework. Since students in the current economy use a variety of ways to communicate that are not all written (or typed) text, in what ways could the framework reimagine prompts and tasks that provide students with opportunities to communicate in a variety of different types of "text?"

Editorial Comments

Two of the 21 submissions included specific edits to current language or formatting suggestions. Summary comments follow.

- In Appendix C (NAEP Writing Achievement Level Descriptors [ALDs]), it is difficult to notice differences between levels for each grade. Comparisons across grade levels cannot be made, so including them side-by-side is not helpful. The ALDs can be reformatted to include the descriptions by grade levels rather than across grade levels.

- In Appendix C (NAEP Writing Achievement Level Descriptors [ALDs]), a suggestion was made to organize ALD by criteria for evaluating responses (i.e., Development of Ideas, Organization of Ideas, Language Facilities and Conventions). This change would also

support the rubric language. The table on page 71 can be reformatted to reflect the criteria for evaluating responses, and headings and bulleted items can be created on the ALDs.

- The first communicative purpose could be written as, "To persuade, in order to change the reader's perspective or affect the reader's action."

- Change "point of view" to "perspective." Perspective is a broader term that will not derail students by developing a narrow focus on point of view.

- Appendix B1 (Preliminary Holistic Score Guide for "To Persuade"): This is difficult to read and use holistically so suggest creating a chart by domain. The chart can have one column for development, one column for organization, one column for language/conventions, and six rows (i.e., one for each level). The descriptions will be included in those cells.

- Appendix B2 (Preliminary Holistic Scoring Guide for "To Explain"): For scores 1 and 0 on pages 65–66, there needs to be a more discernible difference between the scores. A score of 1 could be "little explanation of the subject" instead of "little to no explanation of the subject." The description for a score of 0 can remain the same.

- Appendix B2 (Preliminary Holistic Scoring Guide for "To Explain"): Change the word "marginal" in the definition for a score of 2 to "limited" or "minimal." "Marginal" means minimal for requirements; however, a score of 2 does not meet all requirements for a satisfactory response.

- Note that the Chapter 1 Overview may need revisions, depending on revisions made to other chapters.

- In the "Conclusion," while the use of word processing software is widely available, it seems that statements related to students' ability to write to communicate in the 21st century might be better expressed as "young people's ability to use 21st-century tools to compose writing" or something similar.

- Add "email" in addition to or in lieu of typical letters to be more consistent with the electronic environment of assessment.

- The 2017 NAEP Writing Assessment Framework has a lot of information about moving from a paper-pencil to a computer-based writing assessment. It is no longer necessary to justify the use of computers to assess student writing.

Update Research

Two submissions recommended updating the research in the framework. Specific areas to update included cultural responsiveness, current practices in writing instruction, student performance, and statistics. Summary comments follow.

- The research reference studies are more than 10 years old.

- The current research base of the NAEP frameworks is dated and requires a refresh. Updated research should include a focus on cultural responsiveness to ensure that prompts provide a wide array of cultural context and reflect multiple student groups. Also, consider updating the research base of NAEP frameworks to include more information about current practices in writing instruction and student performance.

- Incorporate updated statistical data and findings from previously investigated research questions. This includes updating framework references to reflect revisions in policies and other sources, such as the WPA Outcomes Statement for First-Year Composition.

- Reviewers should evaluate the accuracy of earlier statements. For example, is more known now about the "impact for form" than in the 2017 release? Do recent developments in and the ubiquity of software tools change the nature of the "commonly available tools" that students are permitted to use in composing on the computer?

- Include findings from the study on the Grade 4 computer-based writing assignment.

- As inclusion and justice become more prominent in educational assessment, new research on topics such as cultural and linguistic diversity, asset framing, and access to technology and information must be considered in the theory, construct, and assessment of writing. This research has significant implications for large-scale assessments of writing, particularly regarding multilingual students, and the use of monolinguistic competencies, the elimination of deficit-thinking in rubric performance level descriptors, and the need for greater diversity in education assessment professionals.

Other Comments/Questions

Finally, there were six additional submissions, including suggestions and comments, regarding the NAEP Writing Assessment Framework. These ranged from adding assessments to asking questions about how student creativity is factored into the scoring rubrics. One comment also focused on how instructional practices merge reading and writing so perhaps the construct should be changed from writing to English language arts. Summary comments follow.

- Include an assessment for writing in arts (e.g., literary, folk, song).

- Include diverse stakeholders who bring expert cultural and linguistic representation to the framework design committees. This should occur through tiered committees to provide diverse and inclusive perspectives representative of the students being assessed.

- We are curious about the framework's approach to student motivation and engagement, which are key detriments of the effort put in academic tasks, such as assessments.

- What role, if any, does student creativity play in the scoring rubrics (includes creativity in sentence structure, wording, figurative language, etc.)?

- Expand professional learning opportunities on the Writing Assessment Framework to include English language development, English to speakers of other languages, bilingual, and language teachers.

- Current classroom instruction incorporates an intentional merging of reading and writing. More often, writing instruction extends from content that students have read and discussed in a class context. This instructional practice provides NAEP a unique opportunity to merge the Writing Assessment Framework with the Reading Assessment Framework.

Table 1. List of People/Organizations that Submitted Comments

Name	Title and Organization
Connie Anderson	Managing Owner, Grandmaloutunes
Miah Daughtery, Ed.D.	Vice President of Academic Advocacy, NWEA
Janice Dole, Ph.D.	Professor, Utah State University
Dianne Henderson, Ph.D.	Vice President, ACT
Chester E. Finn, Ph.D.	Distinguished Senior Fellow and President Emeritus, Thomas B. Fordham Institute
Aimee J. Jahns, Ph.D.	Retired elementary reading specialist and adjunct professor
Loretta Kane, Ph.D.	Professor, Berkeley City College
Marta Leon, Ph.D.	Senior Instructional Designer, Learning A-Z: A Cambium Learning Group Company
Sue Livingston, Ph.D.	Professor, LaGuardia Community College
Megan Lopez, M.Ed.	Education Specialist, Secondary English Language Arts|WIDA, Utah State Board of Education
Banks Lyons, M.Ed.	Elementary ELA Coordinator, Tennessee Department of Education
Theresa McEntire	Elementary ELA Education Specialist, Utah State Board of Education
Lori Pusateri-Lane, M.S.	English Language Arts/Fine and Performing Arts Consultant, Wyoming Department of Education
Danielle M. Saucier, M.Ed.	Literacy Specialist, Maine Department of Education
Shawn Washington-Clark, Ph.D.	Teacher Specialist, NBCT, Anne Arundel County Public Schools
Karen Yelton-Curtis	English instructor, Fresno High School
Heidi Faust, Ph.D. (TESOL) Margo Gottlieb, Ph.D. (WIDA) Joel Gomez, Ed.D. (Center for Applied Linguistics)	TESOL International Association WIDA The Center for Applied Linguistics
Organizations	
CenterPoint Education Services	
Florida Department of Education, Division of Public Schools	
Florida Department of Education, Office of Assessment	
Reading is Fundamental	

Recommendations for Gradual, More Frequent Updates to NAEP Assessment Frameworks

Background

One of the Governing Board's legislatively mandated responsibilities is to develop assessment objectives for NAEP, which is operationalized through assessment frameworks and test specifications. The National Center for Education Statistics (NCES) uses the frameworks and specifications to develop items and test forms for administering the assessments. The Board exercises its authority to develop and update the NAEP frameworks through its policy on Assessment Framework Development. This policy was recently updated in March 2022, but there has been continued interest in re-examining the current policy to consider whether and how smaller changes to NAEP frameworks might occur on a more frequent basis rather than waiting 10 years (or more) and making larger changes all at once.

In preparation for the May 2022 Assessment Development Committee (ADC) meeting, Assistant Director for Assessment Development Sharyn Rosenberg prepared a paper outlining various questions and considerations that would need to be addressed to pursue this idea. The Committee discussed the paper and supported the Board staff proposal to commission consultant papers on this topic. Through a contract with the Manhattan Strategies Group (MSG) and subcontract with the Human Resources Research Organization (HumRRO), papers were commissioned from six consultants who were intended to represent different perspectives and experiences on this topic:

- Carol Jago, former Governing Board member and ADC Chair
- Andrew Ho, former Governing Board member and Chair of the Committee on Standards, Design and Methodology (COSDAM)
- Jessica Baghian, former state leader in Louisiana
- Stanley Rabinowitz, psychometrician with extensive experience working on state assessments and the national exams in Australia
- Ada Woo, psychometrician with extensive experience working on certification exams
- Alicia Alonzo, former member of the NAEP Science Standing Committee, and the committee that recently updated the 2023 TIMSS Science Framework using a process similar to what has been proposed for updating NAEP assessment frameworks

Independent of the consultant papers commissioned by Board staff, Lorrie Shepard of the NAEP Validity Studies (NVS) Panel wrote a comprehensive white paper on this topic, and it was published on the NVS website in November 2022. All papers were included in the November 2022 ADC materials. Each expert prepared a few PowerPoint slides summarizing the most salient points of their papers; those slides are included in this attachment.

Board staff organized a virtual technical panel meeting on January 31, 2023 with the paper authors to discuss ideas raised for the purpose of informing recommendations for how to proceed with the Board policy and procedures for updating NAEP frameworks. Minutes summarizing the technical panel meeting were shared with ADC members shortly before the March quarterly meeting and are included in this attachment.

During the March ADC meeting, Sharyn Rosenberg presented key takeaways from the technical panel meeting and ADC members provided input about next steps, including pros and cons of trying out a new framework development approach with various subjects.

Proposed Changes to Structure and Substance of Assessment Framework Development Policy

In the current NAEP framework development process, a contract to update a framework is not awarded and panels of experts are not convened until the Board formally adopts a charge to launch the update. This means that the initial work to inform whether an update is needed, and if so, the intended scope of that update, is performed on a more ad-hoc basis. This pre-work may include performing research studies and/or other analyses of state trends in education and assessment, collecting public comment, commissioning consultant papers – and synthesizing what implications the lessons learned from these efforts have for a Board charge that provides initial policy guidance and direction for a framework update project.

The original impetus for considering changes to the Board's framework development policy was to make it easier to maintain trends if smaller changes were made on a more frequent basis (this idea is sometimes referred to as "incremental frameworks"). In order for it to be feasible to make smaller changes to frameworks on a more regular basis, it is necessary to have a standing group of content experts monitor the state of the field in a given subject area and its implications for the NAEP assessment framework. As Lorrie Shepard articulated in her NVS paper, "Experts already familiar with NAEP's purposes and structures would monitor evidence from the field and propose needed framework changes to NAGB" (p. 35).

The ongoing efforts to explore the feasibility of changing the NAEP framework process have clarified that it may be of great value to use standing groups of NAEP content advisors to implement a more coherent, systematic process of monitoring changes in a field and advising the Board on whether and how to update NAEP assessment frameworks, both in anticipation of and in response to various changes in a field. Such an approach could be used to achieve many goals, including but not limited to the Board's strong desire to avoid breaking the trend lines for the reading and mathematics assessments in particular. A standing group of NAEP content advisors could also serve other purposes, such as helping the Board navigate contentious issues and debates in NAEP subject areas. Implementing a formal framework monitoring process would be a major change to the current NAEP framework process and would require specification of a principle in the policy statement.

Another major change to the structure of the current framework policy that is implied by recent conversations of proposed changes is the need to articulate and differentiate processes for minor updates to frameworks from processes for major updates to frameworks. The current policy includes a passing reference to the possibility of minor updates but does not describe what process should be used to implement them.

The current policy would need to be revised to in the following ways to support the changes proposed:

- Establishment of a new principle to formalize the framework monitoring process via a standing group of NAEP content advisors in each subject area (tentatively called a "Content Advisory Group")
- A need for clear distinction between "minor changes" and "major changes" to frameworks
- Establishment of a new principle intended to describe guidelines for implementing minor changes to frameworks
- Elimination of references to a Steering Panel (even for major changes to frameworks) given the proposal that members of a Content Advisory Group would work with a Development Panel to implement more extensive changes

The structure of a revised policy, along with needed edits and clarifications, is described below.

Introduction (largely unchanged):

- Need to decide whether this section should state upfront that the primary goal for updates to reading and mathematics frameworks is to avoid breaking the trendlines

- Need to consider whether and how the list of involved stakeholders applies to minor changes to frameworks

Principle 1: Elements of Frameworks (unchanged)

Principle 2: Framework Monitoring Process (extensive changes)

- This would be a new principle that contains some of the information currently in principle 3 (Periodic Framework Review) but restructures it to reflect a more formal monitoring process using a Content Advisory Group.

- *Guideline 2a* would define Content Advisory Group, including the structure and composition (approximately 8-12 members per subject area, at least some with previous NAEP experience and also mix of content/policy expertise in the given subject area, how to keep the group consistent enough but still allow for new perspectives, etc); function of group (stay current on implementation of framework/item development, monitor changes in the field with potential impacts for NAEP frameworks, suggest research studies to be conducted and other relevant information to be collected/synthesized, provide direction to Development Panel when necessary to convene for major updates, etc). For reading/math in particular, the explicit goal is to minimize disruption to trend. For other subjects, maintaining trend is a goal when possible.

- *Guideline 2b* would describe the process and frequency (Every administration? Every 4 years? More often for Reading/Math than other subjects?) for a Content Advisory Group to formally review a NAEP framework and make a recommendation to the ADC about

whether no revisions, minor revisions, or major revisions are advisable. Need to define minor vs. major revisions (tentatively, minor revisions seem like they could include clarifications in the text of a framework document that do not have direct implications for definition of the construct, such as updated references or addition of a new item type, and/or small changes to the construct where most of the construct is unchanged). Recommendation from Content Advisory Group may be based on relevant research and external input.

- *Guideline 2c* would indicate that the ADC prepares a recommendation for full Board approval. Minor updates may be carried out directly by the Content Advisory Group with additional contributors if desirable whereas major updates shall include the convening of a Development Panel. Framework revisions are subject to full Board approval.

- *Guideline 2d* would indicate that Board discussion of a recommendation from a Content Advisory Group on whether to proceed with no change, minor updates, or major updates to a NAEP assessment framework should include major policy and assessment issues in the content area. Such issues shall be identified through the guidelines described in Principles 2a and 2b).

- *Guideline 2e* (based on current guideline 2b): With consideration of the policy and assessment issues in the content area, the Board shall develop a charge to articulate the need for an update to the framework and to specify the scope of the update, policy guidance, constraints (including but not limited to those imposed by the NAEP legislation), and any specific tensions to resolve in the development of framework recommendations. The Board charge shall explicitly address whether maintaining trends with assessment results from the previous framework should be prioritized above other factors (for reading and mathematics in particular, this is expected to be the case).

Principle 3: Development and Update Process for Minor Changes (new)

There is a need for a new principle to articulate a scaled back process for minor changes (the current policy is silent about how minor changes would be implemented). This principle should indicate: when/how others might be involved in addition to Content Advisory Group; what role (if any) for a Technical Advisory Committee (TAC) and/or additional practitioners; whether/when to conduct public comment or seek other forms of external input on proposed changes.

Principle 4: Development and Update Process for Major Changes (largely unchanged from the current Principle 2)

- The current guidelines 2a and 2b are no longer necessary here given the proposal to shift this information to the new Principle 2.

- References to the Steering Panel should be removed. This principle would need to articulate how the Content Advisory Group is intended to interact with and/or provide oversight of the Development Panel for major updates (for example, a few members of the Content Advisory Group could serve on the Development Panel in leadership positions).

Principle 5: Elements of Specifications (unchanged from the current Principle 4)

Principle 6: Role of the Governing Board (largely unchanged from the current Principle 5)

- Some edits would be needed to eliminate references to the Steering Panel and clarify distinctions between minor and major changes to frameworks.

During the upcoming May meeting, ADC will discuss the proposed changes to the framework policy. If the Committee agrees with the general direction of the proposed changes, staff will work on developing specific edits for Committee discussion at the next Board meeting.

ADC may also consider the potential merits and drawbacks of convening a Content Advisory Group for one or more subject areas to provide direction to ADC and Board staff on how to prepare for potential updates to frameworks. The current policy on framework development is silent on how the "pre-work" is conducted, so it would be possible to put together a Content Advisory Group to provide input upfront without fully committing to revising the policy of how changes to frameworks are implemented.

NAEP Framework Development Reaction Paper

Carol Jago

1

Who should decide when revisions to a framework are needed?

The ultimate decision, of course, rests with the National Assessment Governing Board, but I believe NAEP standing committees are ideally positioned to recommend to the board when developments in the field and/or in assessment technology necessitate incremental framework updates.

These standing committees, made up of individuals not only knowledgeable in their fields but also deeply knowledgeable about the NAEP assessments, could also be charged with identifying and recommending to NAGB when disruption in the field is so great that a full-scale revision process is needed.

While it is important not to be distracted by every shiny new thing, to retain their place as authoritative measures of student performance, NAEP assessments must be relevant.

2

Exploring a process for updating frameworks more often with smaller changes

Updating frameworks more often with incremental revisions should be more efficient, but NAGB will need to be careful how changes to policy are communicated to the public. We don't want it to appear that the process is becoming less transparent. It might be a good idea to explore how PISA and TIMMS have handled this issue. Currently the process for NAEP framework development includes extended periods of time for public comment. Inevitably this step slows down the process of keeping frameworks updated to glacial speed.

Taking a more positive spin on this issue, NAGB could publicize the smaller changes to an adopted framework in succinct, "Good News!" updates. It will be important to make clear to all that the ultimate purpose of any change in a framework is to improve the assessment thereby improving education.

3

Anticipating potential unintended consequences

The "debates" in reading and mathematics never really end; they only settle down for short periods of truce and then resurface with renewed vehemence. NAEP results play a role in this endless tug-of-war, particularly when student performance is disappointing. The pendulum metaphor is clichéd but apt.

Alas, we are likely to be revisiting certain issues again and again. Nothing is "settled" for long.

NAEP frameworks are currently voluminous documents. What if they were reconceived as much less detailed guidelines for a national assessment? More like a roadmap for item development than a description of the field.

Currently NAEP frameworks read something like national pronouncements.

Maybe NAEP frameworks try to do too much.

4

Slide 1

Measuring change in a changing world:
Toward efficient measurement of aggregate educational progress

link
link

Andrew Ho, *Harvard Graduate School of Education*
Summary of a paper available <u>here</u>.

1

Slide 2

My Recommendations

1. Only task framework panels for new subjects or rarely administered subjects that require a relaunch;

2. For all other subjects, create (or revise the charge of) standing framework committees to advise the Board and consult with the National Center for Education Statistics (NCES) on necessary incremental changes to existing frameworks and specifications; and

3. Adopt different perspectives on trend reporting and validation, including
   a) a "moving window" perspective on trend validation,
   b) three different levels of "bridge studies," and
   c) differences in validation for developing an index vs. a scale.

2

Slide 3

1. Reserve framework panels for subjects that are new or require relaunch

- Current framework panels are better suited for revolution than evolution. Panelists are often motivated toward substantial revision.

- Framework panels remain a useful tool but should be reserved for new subjects.

- The longer the existing trendline and the more granular (state- and district-level) the aggregation for reporting, the more important it is to avoid a new framework panel and rely instead on standing committees tasked with incremental adaptation.

- I do not believe that the Reading and Mathematics frameworks should ever be relaunched in a manner that suggests a sudden and discontinuous "new Reading" or "new Mathematics."

3

Slide 4

2. Create (or revise the charge of) standing framework committees to update frameworks for existing subjects incrementally

- Standing committees should meet regularly with NAGB and advise on or direct framework revisions for every administration of their assessment.

- These revisions may be prospective due to extended timelines for implementation given NCES constraints on item and task development and field testing.

- Membership terms that overlap and rotate, like Governing Board terms, can preserve institutional knowledge.

- Standing committees can also serve a useful bridging role between the Governing Board and NCES that deepens coordination and communication between NAEP governance and NAEP operations.

4

Slide 5

3. Adopt different perspectives on trend reporting and validation

a) "The bamboo that bends is stronger than the oak that resists."

b) Three bridge studies

c) Can we consider NAEP an Index?
Left: A market basket of goods.
Right: NAEP Math Subscale Trends

5

1

2

3

4
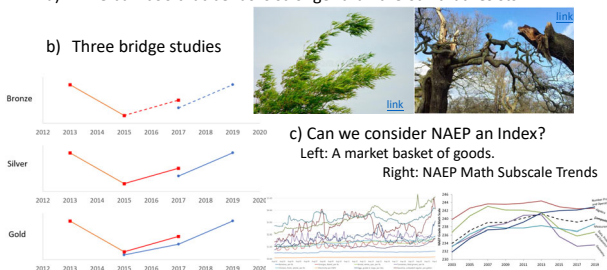
5

NAEP Framework Development:
Jessica Baghian Response

1

Overview

State leaders rely on NAEP to (1) ensure the rigor of state-defined proficiency benchmarks, (2) compare their ranking relative to peer states, and (3) understand their state's trends in comparison to the rest of the nation.

The National Assessment Governing Board (Governing Board) is contemplating updating the NAEP assessment frameworks on a more frequent basis. The Governing Board's motivation includes cost; relevance; operational adjustments based on lessons learned; and smaller, frequent changes instead of infrequent, larger changes.

American education does not evolve quickly enough to necessitate changing the NAEP assessment framework more often than once every ten years. Allowing more frequent reconsideration will almost certainly lead to more changes – and every such change increases the risk to the trend.

Recommendations:

- Prioritize maintaining stable trend lines and, therefore, review the framework only once every 10 years.

- When frameworks are updated at the ten-year mark, gradual operationalization should be allowed. The timeline for such operationalization should be set at that time.

2

Recommendation Rationales and J. Baghian Responses

Proposed NAEP Rationales	J. Baghian Response
Reducing costs	It is unlikely that allowing more frequent changes will reduce cost. More change translates to more item creation, more committees, more standard setting, and more reporting adjustments. Regardless, this is an insufficient justification for risking the ability to compare results across states and over time.
Increasing relevance	As a state policymaker, NAEP was relevant for many reasons, but it did not dictate the academic standards and content learned by my state's children. Annual framework reconsiderations and related tweaks, as are made in many state assessments, are not necessary for NAEP. In fact, they risk NAEP's greatest value – the trend line and the valid state comparisons.
Adjusting to lessons learned	If there is a serious flaw or issue with the test design, test items, etc., responsible test makers should absolutely respond. However, operationalizing the framework (e.g., writing items, data analysis) is different than creating the framework. The process should continue to allow for reasonable technical adjustments in alignment with the Board's overall framework directives.
Smaller changes over time to minimize the risk to the trend.	When a framework is reconsidered each decade, part of the consideration should be – how different is the updated framework from the previous framework, and what is the safest way to transition while maintaining trend? The plan should reflect a "decide once" principle here. Change the substance and determine the operational strategy in tandem. This approach allows for gradual adjustment without revisiting and re-questioning the framework every few years – a practice that assuredly risks the trend.

3

Slide 1

3. Adopt different perspectives on trend reporting and validation

a) "The bamboo that bends is stronger than the oak that resists."

b) Three bridge studies

link

link

Bronze

2012 2013 2014 2015 2016 2017 2018 2019 2020

Silver

2012 2013 2014 2015 2016 2017 2018 2019 2020

Gold

2012 2013 2014 2015 2016 2017 2018 2019 2020

c) Can we consider NAEP an Index?
Left: A market basket of goods.
Right: NAEP Math Subscale Trends

1

Slide 2

EDMETRIC

Keeping NAEP Relevant:
Considerations for Smaller, More Frequent Changes
to NAEP Assessment Frameworks

Stanley Rabinowitz, Ph.D.
EdMetric LLC

2

Slide 3

EDMETRIC

❖ **Considerations, Trade Offs, and Competing Priorities**

- What role does NAEP want to be play?

- What events require a new/revised framework?

- How does NAEP operationalize the reality that "Nothing is ever perfect"?

- Are there inherent differences within and across content areas?

- What constitutes a "change"?

- When does devotion to trend work against the interests of NAEP?

3

Slide 4

EDMETRIC

❖ **Considerations, Trade Offs, and Competing Priorities (cont.)**

- What is the impact of framework revisions on NAEP's validity and equity?
- Is the framework process (the NAEP "way") healthy?
- How can NAEP avoid fads (or hoopla)?
- How can NAEP balance dollar costs versus opportunity costs?
- What image does the Governing Board wish to project: Microsoft versus Apple?

Final Thoughts

In conclusion, the review of the NAEP framework development and revision policy and processes is timely, necessary, and complex. The Governing Board should begin by determining whether its goal is to update current practice or create a new model. The debate should focus not just on the pros and cons of various approaches, but on the likelihood that unintended, unanticipated consequences will compete with expected enhancements.

4

NAEP Assessment Framework Update: Lessons from Certification and Licensure Testing

Ada Woo, PhD
Ascend Learning
January 31, 2023

1

Continuously Assess the Relevance of Assessment Frameworks

- Certification and licensure testing programs conduct practice analyses to ascertain the KSA needed to perform competently in a particular profession.
- Rapid evolution in technology and practice accelerated changes in many professions. These changes often led to more frequent or continuous practice analysis studies.
- For example, the **Federation of State Boards of Physical Therapists** have conducted practice analyses every five years historically. Starting in 2018, the FSBPT analyzed practice annually in addition to conducting its regular practice analyses.
- Results of these continuous practice analyses may serve as leading indicator of change. They also help the testing organizations to remain proactive in their framework development.

2

Emerging Knowledge and Expansion of Content Domain

- The **National Council of State Boards of Nursing** is expanding its licensure exam to include clinical judgment and decision making. The process began with a strategic practice analysis in 2013, leading to a new nursing clinical judgment framework. New items were field tested in 2017. NCSBN used the data to inform its new scoring model and administration plan. The new exam and test plan will start in April 2023.
- The updated assessment framework includes an additional three case studies (18 items) on each exam and a range of standalone clinical judgement items. The new exam will be a mix of traditional and clinical judgment items, with clinical judgment items not exceeding 20% of the test.
- The **Association of International Certified Professional Accountants** is undergoing similar assessment framework changes. Informed by its 2020 practice analysis, the updated exam will include a core component and three disciplines. All examinees will be required to take the core and select one of the three disciplines (tax compliance and planning, business analysis and reporting, or information systems and controls).

3

Develop a Consistent Assessment Framework Across Multiple Subject Areas

- The **National Board for Professional Teaching Standards** develops assessments in over 30 certificate areas along learner ages and subjects (e.g., Early Adolescence English Language Arts and Middle Childhood Generalist). While specific standards are developed for each certificate area, the NBPTS follows the same assessment framework for all certifications.
- All NBPTS certificate assessments begin with the Five Core Propositions. The five core propositions articulated the vision for accomplished teaching, the construct on which the NBPTS certifications are based. From these five core propositions, NBPTS developed standards specific to each certificate area.
- NBPTS also uses the same assessment formats and scoring design across all certifications. All certifications contain four parts. Each part is assessed with different formats, ranging from multiple choice questions to video portfolios. The same scoring design is used across all certification areas, allowing for trend monitoring and comparisons both within a certification area and across multiple areas.
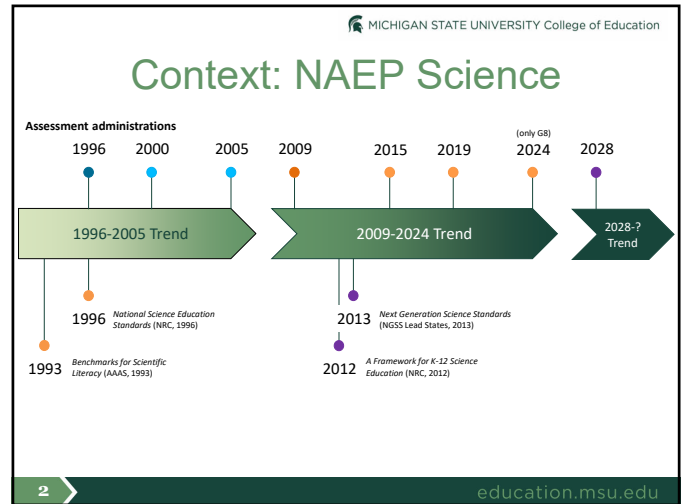
4

Slide 1

MICHIGAN STATE UNIVERSITY College of Education

NAEP Science as a Context to Consider Options for NAEP Framework Revision

Alicia C. Alonzo

education.msu.edu

1

Slide 2

MICHIGAN STATE UNIVERSITY College of Education

Context: NAEP Science

Assessment administrations

1996 2000 2005 2009 2015 2019 2024 (only G8) 2028

1996-2005 Trend   2009-2024 Trend   2028-? Trend

1996 — *National Science Education Standards* (NRC, 1996)

2013 — *Next Generation Science Standards* (NGSS Lead States, 2013)

1993 — *Benchmarks for Scientific Literacy* (AAAS, 1993)

2012 — *A Framework for K-12 Science Education* (NRC, 2012)

education.msu.edu

2

Slide 3

MICHIGAN STATE UNIVERSITY College of Education

Less frequent, larger framework shifts

Prioritizing current (& future) standards

- Framework shifts made "immediately" to reflect new goals for student learning (e.g., PISA)
- Can this process be more efficient?
  - Rely more on consensus processes used to develop standards?
    - However… there are still decisions to be made, including how to deal with variation in state standards
  - Limit revisions to those needed to reflect new standards?

Prioritizing current practice

- Framework shifts made later to allow for shifts in practice & time to address new assessment challenges
- What triggers these shifts?
  - How can shifts in practice be monitored to know when a large framework shift is appropriate?

education.msu.edu

3

Slide 4

MICHIGAN STATE UNIVERSITY College of Education

Smaller, more frequent shifts

- To frameworks
  - Utility
    - Useful if prioritizing current practice
      - Responsive to current practice as stakeholders adjust to new expectations introduced more suddenly
    - More responsive to advances in assessment technology
    - But… not appropriate to capture shifts in the field before consensus has been reached
  - Timing
    - Unclear when framework shifts are appropriate (especially given slow pace of changes in practice)
    - A fixed schedule of revisions (e.g., with every administration) could lead to unnecessary changes
- To framework implementation
  - Framework revisions could be responsive to new consensus in the field, while acknowledging that changes in practice occur incrementally—more natural balance between standards and practice
  - Ability to be responsive to new assessment knowledge and to speed of changes in practice
  - Could reflect gradual progress towards ambitious goals (i.e., providing a forward-looking representation of goals for student learning, so that status quo is not signaled as satisfactory)

THE BOTTOM LINE

Clearer guidance is needed about how NAEP seeks to balance between current standards and current practice

education.msu.edu

4

Slide 1

NAGB: NAEP FRAMEWORKS EXPERT PANEL | JANUARY 31, 2023

White Paper:
NAEP Framework and
Trend Considerations

Lorrie Shepard
University of Colorado Boulder

MAKING RESEARCH RELEVANT

AMERICAN INSTITUTES FOR RESEARCH | AIR.ORG

Slide 2

Recommendations for an "Evolutionary Approach" to Framework Revisions

Points of Agreement with other experts including NASEM (2022)

- NAGB should develop a more explicit policy to protect trend and, at the same time, ensure the relevance of construct representation by providing for ongoing, incremental revisions to frameworks.
- Standing subject-matter committees should have greater responsibility to ensure continuity and integration across stages of the assessment development process and to make recommendations for gradual framework revisions.

Additional Important Points

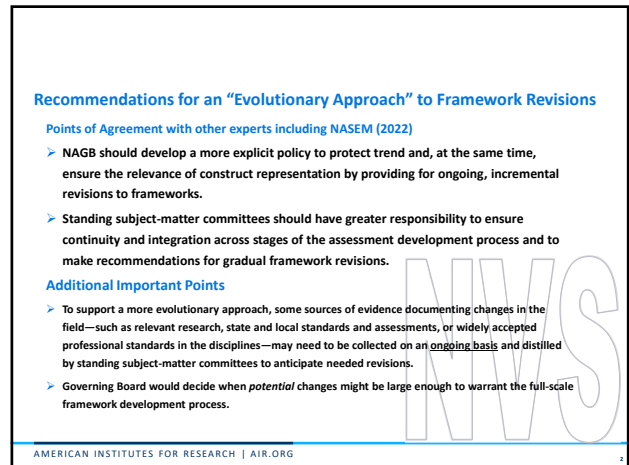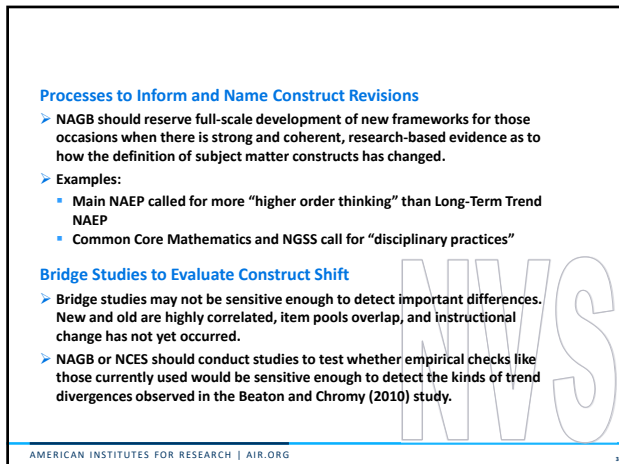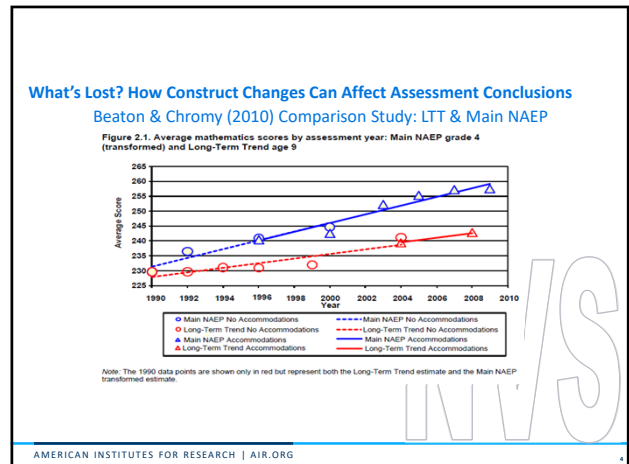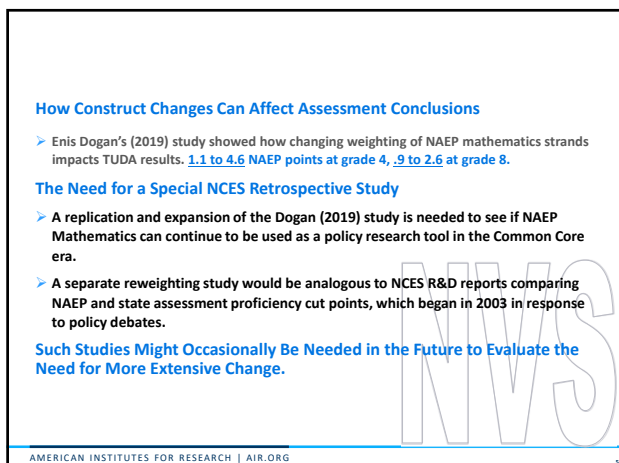- To support a more evolutionary approach, some sources of evidence documenting changes in the field—such as relevant research, state and local standards and assessments, or widely accepted professional standards in the disciplines—may need to be collected on an ongoing basis and distilled by standing subject-matter committees to anticipate needed revisions.
- Governing Board would decide when *potential* changes might be large enough to warrant the full-scale framework development process.

AMERICAN INSTITUTES FOR RESEARCH | AIR.ORG

Slide 3

Processes to Inform and Name Construct Revisions

- NAGB should reserve full-scale development of new frameworks for those occasions when there is strong and coherent, research-based evidence as to how the definition of subject matter constructs has changed.
- Examples:
  - Main NAEP called for more "higher order thinking" than Long-Term Trend NAEP
  - Common Core Mathematics and NGSS call for "disciplinary practices"

Bridge Studies to Evaluate Construct Shift

- Bridge studies may not be sensitive enough to detect important differences. New and old are highly correlated, item pools overlap, and instructional change has not yet occurred.
- NAGB or NCES should conduct studies to test whether empirical checks like those currently used would be sensitive enough to detect the kinds of trend divergences observed in the Beaton and Chromy (2010) study.

AMERICAN INSTITUTES FOR RESEARCH | AIR.ORG

Slide 4

What's Lost? How Construct Changes Can Affect Assessment Conclusions
Beaton & Chromy (2010) Comparison Study: LTT & Main NAEP

Figure 2.1. Average mathematics scores by assessment year: Main NAEP grade 4 (transformed) and Long-Term Trend age 9

Note: The 1990 data points are shown only in red but represent both the Long-Term Trend estimate and the Main NAEP transformed estimate.

AMERICAN INSTITUTES FOR RESEARCH | AIR.ORG

Slide 5

How Construct Changes Can Affect Assessment Conclusions

- Enis Dogan's (2019) study showed how changing weighting of NAEP mathematics strands impacts TUDA results. 1.1 to 4.6 NAEP points at grade 4, .9 to 2.6 at grade 8.

The Need for a Special NCES Retrospective Study

- A replication and expansion of the Dogan (2019) study is needed to see if NAEP Mathematics can continue to be used as a policy research tool in the Common Core era.
- A separate reweighting study would be analogous to NCES R&D reports comparing NAEP and state assessment proficiency cut points, which began in 2003 in response to policy debates.

Such Studies Might Occasionally Be Needed in the Future to Evaluate the Need for More Extensive Change.

AMERICAN INSTITUTES FOR RESEARCH | AIR.ORG

Expert Panel Discussion on NAEP Framework Processes

Meeting Minutes of January 31, 2023

<u>Experts Present</u>

Alicia Alonzo
Jessica Baghian
Andrew Ho
Carol Jago
Stanley Rabinowitz
Lorrie Shepard
Ada Woo

<u>National Assessment Governing Board Staff</u>

Lesley Muldoon
Rebecca Norman Dvorak
Sharyn Rosenberg

<u>Human Resources Research Organization (HumRRO)</u>

Emily Dickinson
Sheila Schultz

<u>Manhattan Strategy Group (MSG)</u>

Tara Donahue
Lori Meyer

<u>National Center for Education Statistics (NCES)</u>

Enis Dogan
Dana Kelly
Daniel McGrath
Nadia McLaughlin
Holly Spurlock

In advance of the meeting, a background document with discussion questions prepared by Sharyn Rosenberg of the Governing Board staff was sent to all attendees. That background is included it italics at the beginning of each discussion topic.

Welcome and Overview

Sharyn Rosenberg, Assistant Director for Assessment Development, welcomed the experts and other attendees to the meeting. She began the meeting by providing background on NAEP to provide context for the discussion topics. She highlighted that maintaining trend is important to the Board and many other stakeholders, particularly for math and reading. The current process of updating NAEP frameworks is perceived by some as causing unnecessary threats to trend. The current process also is increasingly difficult to implement, time-consuming, and expensive. Rosenberg noted several important points that must be considered if changes are made to the framework development process, including (a) NAEP must continue to be perceived as relevant, (b) some tension exists between implementing an abbreviated process and gathering sufficient input from external stakeholders, (c) attention must be given to potential unintended consequences, and (d) much work will be required to work through implementation details and potential challenges if changes are made to the process. Rosenberg briefly reviewed the discussion topics before proceeding to have experts share their input on specific questions that were sent to participants in advance of the meeting.

Proposed Role and Function of Subject-Specific NAEP Content Advisory Groups

Several people have suggested that the NAEP Standing Committees (whose current scope is strictly to review NAEP items and are constituted under contract to the NCES item development contractor) be used to recommend framework changes, and in some cases, to also carry out recommended updates. The rationale for this recommendation was explicitly articulated by Lorrie Shepard: "Experts already familiar with NAEP's purposes and structures would monitor evidence from the field and propose needed framework changes to NAGB."

Based on some existing constraints (including contractual issues, areas of oversight, and the amount of time required), it may not be feasible to implement this recommendation as stated, but it would be very useful to discuss how to address the spirit of this recommendation. For example, it may be possible to constitute a standing content advisory group consisting of the Standing Committee members who are rotating off the committee, along with former Assessment Development Committee (ADC) members who have expertise in the content area, and/or participants in recent NAEP framework updates.

- *What would be the intended roles of a NAEP content advisory group, in addition to regularly reviewing a framework and making recommendations to the Board about necessary updates? For example, how might this group help the Board monitor important changes in the field?*
- *What background and previous experiences are needed for participants to effectively serve on a NAEP content advisory group?*

- *Are there recommendations for the structure and function of such a group, such as how many individuals should be included, their roles, how often should they meet, and how to best use their time?*

To begin the discussion, Rosenberg posed several matters for the experts to consider, including the potential roles of a new NAEP Content Advisory Group beyond regular reviews of and recommendations for a framework, the required background and experience of NAEP Content Advisory Group members, and the optimal structure and function of a NAEP Content Advisory Group. She noted trying to move away from the term "NAEP Standing Committee."

Carol Jago agreed that the group's name should not contain "Standing Committee." Andrew Ho noted that he would like the advisory group and its norms and shared values to prevail. Lorrie Shepard echoed Ho's statement, noting that content advisory groups should be "standing," meaning on-going, and members should know the history of NAEP and the specific content area. Ho pondered if including "content" in the group name would be too narrow, proposing the alternative "Framework Advisory Group." He stated that the group should not only have expertise in the content area, but also technical and possibly political expertise, and include members who can have their "ears to the ground" and confer "field legitimacy" to the group. Rosenberg noted she is not attached to a particular name.

Rosenberg asked if technical experts should participate in multiple content-specific groups. Stanley Rabinowitz stated that he liked the idea of technical experts working across content areas, noting that technical expertise may need to be applied differently to different content areas. Alicia Alonzo agreed that she liked the idea of a core group of people focused on content who could consult with a technical expert as needed. Ada Woo added that ideally the group would consist of people with complimentary skill sets (e.g., psychometricians with a policy background) and would benefit from including insiders who could provide a historical perspective.

Rosenberg asked for estimates of how many people would be manageable for the group to include. Jago inquired if group members would serve for a limited term and noted that there are several variables to consider. She suggested that there should be overlapping membership among the groups so that no group ever consists of all new members. She estimated that it would be difficult to pick only 10 members, but more than 20 would likely be difficult to manage. Woo stated that she thought a group of 10 would be appropriate, but that if there is much work to do, it may be useful to identify two separate groups of 10 with a similar makeup of experience and expertise who could work in parallel and meet periodically. Shepard noted that some previous efforts to update NAEP frameworks began by seeking input from five experts and that 10 members is a big step up from that. She added that a group of eight to 10 members would be feasible, with at least half the members having a deep familiarity with NAEP. Rabinowitz stated that about 12 members would be appropriate, and Woo agreed with most of the group that the eight to 12 members would be ideal. Rabinowitz noted that members should have a relevant job, allowing the Governing Board to learn how things are being implemented in the field. Woo

recommended avoiding including too many academicians. Alonzo added that teachers as well as district and state staff would be essential members.

Articulating Existing NAEP Framework Development Policy with NAEP Content Advisory Groups

The ADC recognizes that there may still be a need to occasionally carry out larger scale revisions to frameworks, even if NAEP content advisory groups regularly monitor frameworks. It would be useful to have some discussion about how the use of NAEP content advisory groups might fit into the current framework policy, and what aspects of the current policy may not be necessary when smaller updates are employed.

External outreach and the process of building consensus is a major cornerstone of the NAEP framework development process. Conversely, a shift towards using NAEP content advisory groups exclusively would reduce the timeframe but creates the risk of becoming too insular.

- *Would the NAEP content advisory group potentially replace the need for the initial Steering Panel?*
- *Under what conditions would it be appropriate for the NAEP content advisory group to carry out recommended framework updates (assuming the Board agrees) themselves, versus passing the task off to a more traditional Framework Development Panel or a hybrid group consisting of NAEP content advisors and others?*
- *What types of external input should be sought to support a new process?*
- *How can we ensure that relying more on NAEP content advisory groups for framework updates increases efficiency without making the work too insular?*

Rosenberg asked the experts to reflect on several scenarios, including whether the Board should consider using a Content Advisory Group as a replacement for the larger Steering Panel, even when a larger update is needed, and a Development Panel may be tasked subsequently to develop the recommendations. If so, what rules of thumb should be applied for the Content Advisory Group to handle certain framework changes and pass other changes along to another group to address? How and when would input from external stakeholders be obtained to support this type of process? She noted the answers to these and other questions will help the Governing Board merge important aspects of the current process with creating a new, more efficient process without becoming too insular.

Shepard agreed with Ho's earlier recommendation that the existing full process be implemented only for new content areas, with one amendment—to also implement the full process if a decision is made that the change(s) needed to an existing framework are major. She noted such a decision would likely require some iterations between the Content Advisory Group, the Assessment Development Committee, and the full Governing Board.

Rabinowitz advised a tiered, almost triaged, process that considers the importance of potential changes (e.g., digital NAEP, the introduction of Common Core Standards). The first tier would involve a subcommittee of approximately five members—some educators and some content or

technical experts—who would define the construct and push for an expedited change to the framework. The second tier would require the subcommittee to track all relevant trends (e.g., districts/states, student performance, international) and report on them. Another tier could involve staff reflecting on lessons learned after each NAEP administration and indicating whether that administration might inform them of an evolutionary change or something important that needs to be addressed more immediately. To implement this type of process, an ordering of operations for importance and urgency will be required.

Jago noted that practice related to a content area can sometimes be wide-ranging, with ongoing changes occurring in the field, some of which can occur rapidly (e.g., ChatGPT and writing). She recommended that one role of the Content Advisory Group, or some other responsible professional, should be to objectively look at where the framework currently stands and determine if there is a major difference between it and the field.

Ho shared the opinion that math and reading will never change sufficiently enough to require a relaunch of their frameworks. He stated that the job of the Content Advisory Group should be to protect trends—to ensure the math and reading trends are never broken. For the members to do their job properly, they must maintain current and accurate knowledge of what is happening in the field to prevent a sudden change or recommend bridge studies of different kinds when necessary. They would do this in part by gathering feedback from external stakeholders. Ho also noted that NAEP administrations that are at a higher level of aggregation pose the most risk of not being able to compare the two most recent administrations to each other.

Shepard agreed with Ho's recommendation but noted the importance of a study conducted by Enis Dogan of NCES investigating the impact of reweighting the mathematics subscales to better reflect the weighting of content areas on state assessments. This reweighting impacted the results and interpretations of the mathematics data and may have implications for framework changes. Information from the Dogan study has not been widely shared, and she strongly encourages follow-up special studies to be conducted.

Rabinowitz cautioned the Governing Board to be clear about what is on and off the table before establishing the Content Advisory Group. Members will need to know the rules so they can apply them when making recommendations about whether to change a framework. He noted that everything is important, but that not everything is doable. However, everything is interpretable once you understand the rationale behind the recommendation.

Based on the input, Rosenberg asked the panelists to comment on setting norms or establishing a model to facilitate setting expectations for the Content Advisory Group regarding changes to the NAEP frameworks. She stressed the Governing Board is generally interested in only making changes to frameworks that are necessary and wants to ensure any such changes do not result in more change than is warranted.

Jessica Baghian shared her previous experience, stating that when individuals are asked to review and make a recommendation, they cannot help but suggest many changes they believe

will serve to improve the item, assessment, framework, etc. She expressed doubt that convening a group of experts (regardless of their areas of expertise) will result in only a few or minor suggested changes. In response to Shepard's comment about the Dogan study, she suggested the Content Advisory Group have a role in monitoring what is happening in the field and sharing their findings with states.

Alonzo noted that, although the Trends in International Mathematics and Science Study (TIMSS) employs a consistent model of conducting a review after each administration, TIMSS is inherently more conservative than she believes NAEP should be. She perceives TIMSS to focus on what is currently taught in schools across the participating countries, whereas NAEP is more forward-looking and strikes a better balance between what is taught and what we want students to learn. Changes to TIMSS have been very minor, so Alonzo suggested there may be a way to frame the types and extent of changes to NAEP frameworks that are desired.

Woo added her agreement that changes to NAEP frameworks should be gradual and include only necessary changes. She also agreed it will be difficult to ask group or committee members to review and comment on changes to a NAEP framework and not receive feedback that quite a few changes are needed. She acknowledged that the amount of time between reviews will likely impact recommended changes and incorporating them. She stressed NAEP might adopt a moderating approach that emphasizes communicating to stakeholders that change is coming, why change is needed, and a date when that change will be implemented.

Rosenberg commented that while there is some agreement about having the Content Advisory Group monitor the field and report on necessary changes to NAEP frameworks, she raised some concerns. One worry is whether the Content Advisory Group should have sole responsibility for monitoring what is happening in a field or if it would be better for others to play that role as well to ensure a new development or trend is not missed. She also expressed worry about moving from a process that gathers feedback from external stakeholders at various points during the process to one where only limited external feedback is gathered.

Rabinowitz cautioned that it will be necessary for the Governing Board to not only decide what the new process will be but also how to communicate a new process that exists and plans for transitioning from the old to the new process. This communication will need to describe whether, how, and the frequency to which feedback will be gathered, especially since many external stakeholders will expect to provide such feedback. He agreed the Content Advisory Group should provide input, but he strongly recommended that any new process precludes the Content Advisory Group from making any decisions specifically about content. He believed content decisions should be made by others who understand the broader context of change.

Shepard emphasized the need for the Content Advisory Group to clearly name any change they recommend to the construct and systematically gather input from others. As an example, she suggested obtaining feedback from state assessment directors periodically on what they think is changing in their state and tracking the feedback for reporting back to the Governing Board. She indicated committee meeting minutes will not be sufficient as feedback, suggesting instead

targeted conversations that detail what specifically is happening in the field and why the recommended change must be made in a certain way. She noted overall costs will need to be considered when gathering feedback from external stakeholders, including the public. She suggested thinking about what feedback and comments can be collected cost-effectively every 2 years. She opined that to lead rather than catch up, the Governing Board and NCES should entertain empirical studies (e.g., roll into a Balanced Incomplete Block design that spirals some of the lead state items) or embed tryouts, both of which could provide information similar to the job analysis information gathered for licensure exams to track the need for change.

Potential Meta-Framework for Consistent Elements Across Content Areas

Ada Woo noted that the National Board for Professional Teaching Standards (NBPTS) uses a standardized framework across all certificate areas. This idea was embraced in a recent ADC discussion, where members wondered about the pros and cons of developing a generic framework to specify elements that apply across subject area frameworks, such as the role of equity for NAEP assessment frameworks, universal design elements, item types, etc.

- *What are some pros and cons of developing a meta-framework that could be used as a starting point for all NAEP subject-area frameworks? Is the potential increase to efficiency offset by the need to also continuously monitor and update the meta-framework?*
- *What elements might be included in a meta-framework to make it a worthwhile endeavor? Who would be charged to create the meta-framework? Would it follow the current framework update policy and process?*

To provide context to this discussion question, Rosenberg shared that NAEP frameworks are voluminous documents that tend to serve as a broad pronouncement of the field. She shared the panelists' papers with the Assessment Development Committee (ADC) at the November Board meeting and noted that the ADC embraced a suggestion raised in one paper—the idea of a meta-framework. They were curious to learn more about the pros and cons of such a standardized framework, especially because certain issues are raised when updating a framework regardless of the content area and they were intrigued by how these might be addressed by a meta-framework. She asked the panelists to share their thoughts about how such a meta-framework might work and if creating such a document would result in just one more framework needing to be updated.

Woo shared the example of the National Board for Professional Teaching Standards process, stating they follow the same assessment framework for all certifications. They begin broadly with five core propositions and use the same assessment formats and scoring design across all certifications. She shared that a major advantage of this approach is that a framework exists for each subject area because of the common format and information presented about the core propositions. The major disadvantage of this approach is the difficulty in comparing subject areas because of differences in sample size (e.g., the sample size for English is very large while the sample size for Latin is very small).

Rabinowitz was uncertain about a meta-framework but liked the idea of a meta-committee. He perceived that a meta-committee could develop the rules and processes that various content committees would follow as well as monitor the different content committees.

In writing about the history of NAEP frameworks for the 20th anniversary of the Governing Board, Jago stated that she saw how different they all were and believed it would be easier for readers if the frameworks followed a template. She agreed there could be benefits for framework committees to follow a template instead of reinventing the wheel for common information each framework should include.

Ho stated he would advocate for overarching principles that underlie the frameworks, but he did not agree with thinking about it as a meta-framework. He has advocated for some time that the Governing Board identify core principles and answer such questions as what do we measure and how do we decide on how we measure those elements?

Woo agreed a framework template may be somewhat restrictive; however, she believed stating the core propositions and governing principles will be very important. She noted they can serve as good cost-control measures when dealing with multiple frameworks. Using a template will likely be less expensive than customizing a framework for each content area.

Rosenberg agreed with Jago that the NAEP frameworks are inconsistent in their appearance and the information they contain. She perceived benefits in streamlining the information presented across frameworks even if a template is not applied. Streamlining could include boilerplate-type language such as describing the purpose and background of NAEP, at a minimum.

Shepard agreed with the comment about greater consistency across NAEP subject area frameworks. She also agreed that NAEP frameworks might be doing too much and need not be so elaborate and detailed. She suggested the Content Advisory Group and the frameworks be established in ways that facilitate gathering and summarizing information for the Governing Board and ADC members and that this be done consistently across content areas.

Alonzo indicated having trouble imagining how a meta-framework would look because the subject areas are so different. She liked the idea of defining the process and focusing on what the framework should include. She suggested stakeholders, such as the states, may find it easier to compare information across frameworks simply because they are formatted similarly. She offered that it may not be possible now to identify all information that should be common, but as the frameworks are developed, greater consistency in appearance might, over time, lead to greater consistency of content (as commonalities would emerge naturally by making the frameworks look similar). She cautioned to be flexible when creating a template so as not to be too constraining.

Narrowing the Scope of NAEP Assessment Frameworks

The intended purpose of NAEP assessment frameworks is to (along with the companion Assessment and Item Specification documents) serve as a roadmap for NAEP item development.

Carol Jago has pointed out that NAEP frameworks tend to include a large amount of text describing the state of the field in the given subject area. The generation of this text is a very time consuming and often controversial process. While some limited description of the state of the field may be necessary, it is likely the case that, according to Jago, "NAEP frameworks try to do too much." In contrast to NAEP, for example, the framework documents for the Trends in International Mathematics and Science Standards (TIMSS) are significantly shorter. Reducing the text describing the state of the field would make it much easier to regularly update NAEP frameworks, but it is important to include enough information to avoid compromising the item development process (especially when framework development and item development are carried out through different entities).

- *How might we think about reducing the text of NAEP frameworks without compromising their utility?*

Rosenberg introduced this topic by citing Jago's paper. She acknowledged that while it is necessary for the frameworks to detail what should be assessed, their length is a significant barrier to them being updated more regularly. Rosenberg described the substantial amount of text and in some cases out-of-date citations that must be updated, which requires substantial effort. She further pointed out that much of this text does not have any direct implication for what is on the NAEP assessment. Rosenberg asserted that a simple list of content objectives and practices would not be sufficient and that there needs to be some background and rationale for why these content objectives and practices are to be measured. Rosenberg asked the experts to reflect on the appropriate amount of background and research to be included in the NAEP assessment frameworks and how they could be more consistent across content areas.

Jago stated that the NAEP frameworks are too lengthy and thus not generally read, so they would be more useful if they were more concise. Alonzo added that when using the NAEP Science Framework as part of her work reviewing items on the Standing Committee, they found, even with all the detail in the current framework, it was difficult to determine what is "in" and what is "out." Given that different stakeholders use the framework for different purposes, it might be worth thinking about the item specifications document as containing some of the detail currently in the frameworks themselves (such that the framework does not have to fulfill that purpose). She offered that some elements of the current framework could be moved to an item specifications document.

Rabinowitz noted that more consistency across frameworks would be beneficial because users would not have to relearn how to read the framework when they moved from one content area to another. He also suggested the Governing Board conduct focus groups and surveys to better understand what stakeholders do and do not value about the current frameworks.

Rosenberg asserted that the framework documents would be nimbler if updates to pages of text were not required every time the development process was implemented. She reported that during the current process of updating the 2028 NAEP Science Framework, the panels were told to focus initially on creating an outline of what will drive the assessments rather than spending a lot of time writing a lengthy narrative with citations. After the public comment process, changes

will be made based on the feedback received and the outline will be further fleshed out into more of a coherent narrative. She noted that one possible drawback of this approach is that when the initial framework outline is distributed for public comment, readers might not have enough background or context to understand what is in the framework and why it is included.

Appropriateness of Content Advisory Group for All Content Areas

Based on earlier comments, Rosenberg sought panelists' input about whether the Content Advisory Groups would implement the new framework development process for all content areas or only for math and reading. She worried this could be overly daunting for the Content Advisory Groups and the Governing Board if the process was applied to all seven NAEP subject areas. She asked for thoughts on which subject area might be good to try out the new process— would it be better to try it out on math and reading or a less prominent subject area? She also asked how many subject areas would be feasible for the Content Advisory Groups and Governing Board staff to handle at any given time.

Rabinowitz suggested that the new process can apply to all subject areas, but consideration must be given to available resources. Assuming limited resources, he suggested one option for deciding could be about what is most important—what is more important, preserving trend or being relevant? A second option could be what subject area is most likely to change?

Ho recommended the new process be implemented first for math and reading because the score interpretations and level of aggregation for these two subject areas are particularly important. He offered that all subject areas are important but suggested the level of aggregation and the level of coverage mandated at the state level be prioritized over the perceived inherent importance of a specific subject area.

Although agreeing with Ho about the initial implementation of the new process for math and reading, Woo wondered about the likelihood of the Content Advisory Groups and the Governing Board being able to do so for both math and reading simultaneously. Woo offered there may be less actual work if both subject areas were addressed at the same time, assuming at least some of the comments regarding consistency around the frameworks are implemented.

Rosenberg asked if implementing the new process assumes that the framework has been reasonably updated. She noted the math and reading frameworks have been updated relatively recently, the science framework is currently being updated, and some other frameworks may be updated in the not-too-distant future (i.e., writing, history and civics). She wondered if the new process should be implemented initially for relatively updated frameworks or frameworks that have not been updated in 20 or more years.

Assuming consistency in the framework template, Shepard recommended the process begin with math and reading. However, she stressed the importance of naming the construct change and noted it was good that the Governing Board changed the name of the initial panel from Visioning

Panel to Steering Panel. She suggested following a process like that used by TIMSS where a clear statement is provided for departing from the existing structure.

Woo suggested there be a potential conversation about framing, with the message that the new process is being implemented initially for the subject areas that are most impactful, which likely are math and reading. However, if there is a lot of public sentiment that the most impactful framework is a different subject area, then the framing conversation may need to be different. Regardless of the subject area, she perceived an overall aggregated process implemented in stages across the different subject areas. If an aggregated process is implemented, she suggested there may be less importance placed on the recency of the framework.

Rabinowitz recommended not beginning the implementation of the new process with the math and reading frameworks. He suggested the process could be initially implemented for history because this is an "off to the side" and perhaps a less controversial subject area for which the framework needs updating. He proposed working out the rules and implementing lessons learned before applying the new process to math and reading. He indicated there likely will be a debate about the process, so starting with a subject area like history may facilitate working through the process without also debating issues related to the subject area. Rosenberg pointed out that history may be particularly controversial as a subject area, even though it does not have the same prominence as the NAEP reading and mathematics assessments.

Alonzo advised that if the new process focuses on evolutionary changes, then it does not make sense to start with a subject area that requires a major revision to the framework. However, if the new process focuses on a major revision or relaunch, choosing a subject area where the framework requires major changes may be appropriate. She noted the subject area chosen for the initial implementation should align with whatever in the new process is being emphasized (i.e., evolutionary changes or full process). Ultimately, she recommended starting with a relatively updated framework.

Rosenberg asked the panelists for thoughts about whether each process (regardless of the magnitude of expected change) should begin by having a Content Advisory Group make recommendations regarding how to update the existing framework. That is, even for the frameworks that have not been updated in a long time and may require significant revision, should the process begin with recommendations from a Content Advisory Group (comprised largely of people who have previously worked closely with the NAEP subject) rather than a Steering Panel?

Alonzo agreed that convening such a group to help navigate the framework process may be beneficial, especially in having this group define the changes or parameters for changes to be made. She suggested this group could be a nice test for providing a vision of what the new process might entail, including the potential for how the Content Advisory Group should think about revising the NAEP frameworks.

Ho believed there may be two slightly different charges being discussed and stressed the need to be clear about the charge for this new Content Advisory Group. He perceived the priority for the group should be on implementing evolutionary changes to the frameworks. He preferred the group follow established principles for evolutionary change even in cases when a new framework is launched. He suggested these principles be implemented for math and reading, noting the difference would be that the process would occur more frequently given the aggregation for these two subject areas.

Rabinowitz noted a basic difference between two described goals regarding changes to the frameworks—changes in frameworks to maintain the relevance of NAEP and changes to prevent breaking trend—which appear to conflict. He wondered if the goal is to maintain trend, could that goal apply to math and reading but not to other content areas? He highlighted that there are implications of focusing singularly on each goal, which could result in unintended consequences if applied to only math and reading or applied to all content areas. He suggested the common barrier throughout the discussion has been the divergence in thinking about the process, given these two competing goals. Rosenberg noted, to the extent possible, is the Governing Board desires to accomplish both goals.

Jago noted the common theme when reading the panelists' papers was that the current framework development process is not working and there must be a better process. She cautioned not to get stuck on the details but rather focus on the broader goals of the process, which is that NAEP could lose its recognition in the world if a better framework development process is not identified. Rosenberg acknowledged the original impetus for a new process was driven by maintaining trend; however, she mentioned there are other reasons why the current process is not working well (e.g., costs, long and time-consuming process). She recognized the dilemma of indicating the desire for only "necessary" changes to a framework, especially if it has not been updated in a very long time, and having that statement be received with credibility.

Implications of an Evolutionary Approach to Frameworks for Analysis and Reporting

Separately, Lorrie Shepard and Andrew Ho articulate analytical and reporting implications of an evolutionary approach to framework development, whereby trend comparisons to the most recent previous administrations are more justified than comparisons to the distant past. It would be useful to discuss potential implications for analysis and reporting with such an approach, including the NAEP achievement levels and achievement level descriptions (ALDs).

- *On the Nation's Report Card website, the NAEP results are compared most prominently to the most recent year of administration and the first year of administration. Would an evolutionary approach to framework development suggest that the program should revisit the reporting back to the initial assessment year?*
- *What are other implications of this framework approach for analysis and reporting?*
- *The Board policy on Student Achievement Levels for NAEP specifies that reporting ALDs should be revisited and potentially revised with updates to NAEP frameworks. How would an evolutionary approach to framework development impact the process for updating ALDs?*

Rosenberg described how NAEP results are reported by comparing the current year to the previous year and the first year of administration. She posited that with the framing suggested by Shepard and Ho, it may not be appropriate to compare back to the initial year. She also noted recent work completed to create achievement level descriptors (ALDs) based on content reflected in the current NAEP item pools. Rosenberg asked the group to consider the implications of frequent framework updates for downstream activities, such as analysis and reporting.

Rabinowitz argued that it is appropriate to compare back to the first year until it is decided that the first year is too different from the current year to make meaningful comparisons. He identified three things that would raise concerns about comparing back to the initial administration year: (a) a significant change in NAEP (e.g., digital NAEP), (b) changes in the framework, and (c) changes in the world. Rabinowitz contended that big events, internal or external to NAEP or internal to a particular content area, would determine what the relevant year comparisons are.

Ho noted the potential for misleading data displays if comparisons are not made to an established baseline, but also noted that strict comparisons of current performance to initial performance rely on a lot of pliability. Ho reminded the group that, because NAEP does not report individual student scores or make inferences about individual students, there is little threat of drawing inferences at the aggregate level. He would favor a policy of reporting in a 10- or 12-year moving window, or of reporting for a set number of past administrations.

Jago described her recent experience participating in a project to develop reporting ALDs for NAEP Reading as a facilitator, noting that a shift in the framework would have to be substantial to impact the ALDs. Shepard asserted that the Governing Board should be aware that adding certain kinds of items to the assessment can change the construct being measured and this knowledge can be used to determine if an ALD study is required. She further noted that the weighting of subtests is much more impactful than changes in items.

Final Thoughts

Before closing the meeting, Rosenberg asked the panelists to reflect on the questions raised and share final thoughts or advice.

Shepard reiterated an earlier point she made about bridge studies being insensitive to the kinds of substantive changes that evolve over time. These studies cannot measure or provide the evidence needed to support that a construct is changing. Instead of conducting bridge studies to see if constructs have changed in the short term, she recommended that studies be conducted to evaluate the sensitivity of bridge studies in instances where a change is known, and the research be designed in a way to see whether the change can be detected.

Jago indicated needing to ponder more about the new process, how there might be more in-house roles and responsibilities within the process, and what unintended consequences might result.

More consideration is needed in selecting the eight or 10 members of the Content Advisory Group to determine the appropriate balance of in-house knowledge and external breadth. Factors to consider in membership selection should include real-world experience and the rotation of new and existing members to ensure some overlap.

Based on the papers and meeting discussion, Rabinowitz perceived three competing values that must be considered for the new process—trend, relevance, and equity. If an appropriate balance is not found among these values, he perceived the Governing Board will lose credibility in how people value NAEP. Because something always happens that you do not want, he stated the Governing Board must consider and weigh any unintended negative consequences before implementing the new process.

Woo perceived a major theme throughout the discussion was the importance of preserving trend. If this or another goal is decided, the Governing Board must be clear and upfront about the goal in its charge to the Content Advisory Group. She also pointed out that whatever decisions are made about the new process, someone is going to be unhappy. Documentation and communication about what changes will be made to the frameworks and why will be key for transparency and bringing everyone on board.

Alonzo agreed with the need for the Governing Board to be clear about the purpose of the changes being made. She perceived the tendency will be more toward prioritizing the relevance of NAEP, so if the decision is conservative in making changes to maintain trend, that goal should be clearly stated. She also suggested that defining what changes will be made and the criteria or guidance regarding the evidence needed to support the changes will facilitate the perception of a fair and systematic process.

Baghian stressed the need for the Governing Board to be clear about the purpose and charge for the new process so members of the Content Advisory Group know about any constraints they must work within. She suggested simplifying the process as much as possible and embedding as few opportunities for change as necessary, even if the group has frequent discussions.

Ho also emphasized the need for the Governing Board to be clear and specific about the charge to the Content Advisory Group. He perceived the group's primary role will be to evolve the frameworks in a way that preserves trend. Additional roles for the group include representing the country, schools, and students and how they learn, and making clear recommendations after considering what tradeoffs might be at play. Finally, the group must realize that updating the frameworks will not be a one-time ask, but rather this process will be repeated over and over.

Wrap-up and Next Steps

Rosenberg thanked the experts for their papers and for the discussion, noting that they had given the Governing Board much to think about. She noted that she is chairing a session on this topic in which some of the experts are participating at the April conference of the National Council on Measurement in Education (NCME).

Rosenberg stated that Board staff and the ADC would use the experts' perspectives to help them determine a clear path forward regarding its assessment framework development policies and processes. She noted that this would be the subject of a full Board discussion after the ADC develops some specific recommendations. As the work proceeds, the experts may be contacted for additional input.

IES Institute of
Education Sciences

**Assessment Development Committee
Item Review Schedule
February – December 2023
As of April 25, 2023**

Review Package to Board	Board Comments to NCES	Survey/ Cognitive	Review Task	Approx. Number Items	Status
2/13/2023	3/10/2023	Cognitive	Mathematics (4, 8) *2024 Operational*	Flagged Items Only (4 items)	✓
3/15/2023 (Off-cycle)	4/5/2023 (Off-cycle)	Survey	SQ Reading (4, 8) *2026 Operational (2024 Pilot)*	110-115 pilot and 2024 COVID-19 recovery items*	✓
3/15/2023 (Off-cycle)	4/5/2023 (Off-cycle)	Survey	SQ Math (4, 8) *2026 Operational (2024 Pilot)*	180-185 pilot and 2024 COVID-19 recovery items*	✓
5/3/2023	5/26/2023	Cognitive	Mathematics (4, 8) *2026 Operational (2024 Pilot)*	10 blocks (315 discrete items and 7 SBTs)	
5/3/2023	5/26/2023	Cognitive	Reading Router** (4, 8) *2028 Operational (2024 Pilot)*	94 items	
5/16/2023 (Off-cycle)	6/9/2023 (Off-cycle)	Cognitive	Reading (4, 8) *2026 Operational (2024 Pilot)*	15 blocks (150-162 items)	
7/19/2023	8/11/2023	Cognitive	Reading (4, 8) *2024 Operational*	Flagged Items Only (1 discrete item and 3 SBTs)	
7/19/2023	8/11/2023	Survey	SQ Reading (12) *2028 Existing Item Pool Review*	60-70	
7/19/2023	8/11/2023	Survey	SQ Math (12) *2028 Existing Item Pool Review*	80-90	
9/6/2023 (Off-cycle)	9/29/2023 (Off-cycle)	Cognitive	Mathematics (4, 8, & 12) *2028 Operational (2026 Pilot)*	Concept Sketches (TBD)	
11/1/2023	11/28/2023	Cognitive	Reading (4, 8, & 12) *2028 Operational (2026 Pilot)*	Concept Sketches & Passages (TBD)	

Cross-grade items are included and counted once.
**To support multi-stage testing in 2028.*