



2021 No. 022

Assessment Framework Development Processes

Final Report

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

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

Authors: Will Lorie
Brian Gong
Center for Assessment

Date: March 31, 2021

Assessment Framework Development Processes

Table of Contents

Executive Summary	1
Background and Approach	2
Assessment Frameworks	2
Statement of Work	2
Approach	3
Scope of the Review of Framework Processes	3
Guiding Questions for Review of Framework Processes	4
Definitions	5
Methodology	6
Overview of Methodology	6
Initial Documentation	6
Rationale for Selection of Assessment Programs to Review	6
Additional Documentation Reviewed	6
Organizer: Elements of Framework Processes	7
Key aspects of Framework Processes Relevant to NAEP	10
Descriptions of Assessment Programs Reviewed	11
National Assessments	11
Frameworks for State Assessments	14
International Assessments	17
Professional Standards and Framework Processes	18
Key Findings	19
Implications of NAEP Legislative Mandate for NAEP Framework Processes	19
How this Review Might Inform NAEP Framework Processes	21
Towards Best Practices for Framework Processes	21
Recommendations for Additional Work to Inform Governing Board Considerations	22
References	27
Appendix A: Additional Documents Reviewed	29
Appendix B: Review of Framework Processes in the Pan-Canadian Assessment Program (PCAP)	31

List of Tables

Table 1. Documents Addressing Framework Processes	7
Table 2. Framework Processes Key Components and Questions Addressed by Element	8
Table 3. Key Aspects of Framework Processes Relevant to NAEP	10

Assessment Framework Development Processes

Executive Summary

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

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

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

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

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

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

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

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

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

Background and Approach

Assessment Frameworks

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

Statement of Work

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

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

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

4. Evaluate which practices are appropriate for NAEP's legislative mandates, e.g., curricular-neutrality, pedagogical-neutrality, etc.
5. Describe how current NAEP framework processes reflect or do not reflect these NAEP-appropriate practices.
6. Recommend possible additional work to inform Board considerations.

Approach

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

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

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

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

Scope of the Review of Framework Processes

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

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

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

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

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

Guiding Questions for Review of Framework Processes

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

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

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

7. What are common features of framework processes across all programs, and what appears to be unique to programs or programs with specific characteristics?
8. Which features of framework processes seem most appropriate to those assessment programs with a legislative mandate similar to NAEP?
9. To what extent have NAEP framework processes reflected those features?

Definitions

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

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

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

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

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

Methodology

Overview of Methodology

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

Initial Documentation

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

Rationale for Selection of Assessment Programs to Review

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

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

Additional Documentation Reviewed

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

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

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

Table 1. Documents Addressing Framework Processes

Documents <i>specifying requirements for or reporting elements of framework processes</i>	Documents addressing or potentially addressing standards or best practices
<ul style="list-style-type: none"> • Mandates (Laws, memorandums of understanding, charters, and other agreements – see definitions) • Statements of work • Work plans • Assessment frameworks • Reports • Communiques • Other (websites, presentations, briefs, etc.) 	<ul style="list-style-type: none"> • Standards • Guidelines • Assessment frameworks • Reports • Communiques • Other (websites, presentations, briefs, etc.)

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

Organizer: Elements of Framework Processes

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

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

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

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

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

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

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

Key Aspects of Framework Processes Relevant to NAEP

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

Table 4. Key Aspects of Framework Processes Relevant to NAEP

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

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

Descriptions of Assessment Programs Reviewed

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

National Assessments

National Assessment of Education Progress (NAEP)

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

Initiating Conditions

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

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

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

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

Work Product

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

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

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

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

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

Work Process

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

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

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

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

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

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

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

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

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

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

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

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

Owner, Timeframe, and Approval

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

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

Pan-Canadian Assessment Program (PCAP)

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

The SAT and the ACT

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

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

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

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

Frameworks for State Assessments

Common Core State Standards (CCSS)

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

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

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

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

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

Next Generation Science Standards (NGSS)

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

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

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

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

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

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

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

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

International Assessments

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

Programme for International Student Assessment (PISA)

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

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

Trends in International Mathematics and Science Study (TIMSS)

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

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

Professional Standards and Framework Processes

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

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

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

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

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

A detailed description of the strategies the State used to ensure that its academic content standards adequately specify what students should know and be able to do;

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

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

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

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

Key Findings

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

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

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

Implications of NAEP Legislative Mandate for NAEP Framework Processes

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

Curricular Neutrality

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

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

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

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

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

Diversity of Views

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

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

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

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

Role of Professional Standards

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

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

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

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

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

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

How this Review Might Inform NAEP Framework Processes

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

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

Towards Best Practices for Framework Processes

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

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

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

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

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

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

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

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

Recommendations for Additional Work to Inform Governing Board Considerations

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

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

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

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

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

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

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

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

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

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

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

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

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

1. Content only. The content standard describes what students should know or understand or be able to do but does not include how a student is supposed to demonstrate that knowledge, understanding, or skill.
2. Content with minimal performance descriptions. The content standard includes description of the content and indicates what the student is supposed to be able to do with that knowledge, understanding, or skill. Minimal detail is provided in this performance description. Very many U.S. state content standards use this structure.
3. Content with detailed performance descriptions. The content standard includes description of the content and indicates in some detail what the student is supposed to be able to do with it or how the student is supposed to demonstrate the desired level of expertise. The Next Generation Science Standard’s (NGSS) Performance Expectations (P.E.s) are a widely known example of this approach.

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

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

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

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

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

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

- Determine what is common across the curricula/content standards of the assessed population. An assessment may focus on those things which all curricula agree on; that might be found through a systematic survey of relevant curricula. This is done explicitly for at least one non-U.S. assessment program. (We note that NAEP also has conducted such studies but, to our knowledge, not expressly to test what is common.) Note that the methodology for determining what is common, and assessing whether the process results in something meaningful, is a separate and non-trivial matter that could be addressed ahead of time.
- Refer to education research in the content domain and deliberately ignore curricula/content standards. An assessment may build its content specifications from research only, if available, without referencing curricula. If the research literature is extensive and detailed enough, it may provide sufficient basis to generate content standards, especially if there is broad consensus about the research base. Note: This seems like the least practical to us and the most difficult to specify requirements for. We include it here anyway for completeness.

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

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

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

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

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

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

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix A: Additional Documents Reviewed

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

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

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

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

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

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

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

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

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

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

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

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

**Council of Ministers of Education, Canada (CMEC). (n.d. b). Pan-Canadian assessment program (PCAP) reading assessment framework. (T.Y. 2007). Retrieved from: <https://www.cmec.ca/docs/pcap/pcap2007/Reading-Framework-EN-Apr2013.pdf>

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

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

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

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

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

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

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

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

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

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

Relevance of PCAP

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

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

Assessment Frameworks

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

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

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

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

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

Key Aspects of Framework Processes

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

Authority and/or Legislative Mandate

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

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

Descriptions of Framework Derivation Process

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

Intended Relationship to Academic Standards or Curricula of the Assessed Population

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

Role of Curricula/Content Standards of the Assessed Population

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

Role of Education Research in the Content Area

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

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

Role of Other Frameworks

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

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

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

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

Role of Professional Standards

The PCAP sources do not reference professional standards.

Sources for the Assessment Design

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

Authorship of Framework Documents

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

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

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

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

Constituency Review Processes

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

Processes for Reviewing, Updating, and Revising Existing Frameworks

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

Approval

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