Emerging Ideas for NGSS Assessment

Addendum to the
Science Assessment Item Collaborative
Assessment Framework
for the Next Generation Science Standards

April 2017

THE CENTER ON
**STANDARDS &
ASSESSMENT IMPLEMENTATION**
WestEd CRESST

ACKNOWLEDGEMENTS

ABSTRACT

With support from the Council of Chief State School Officers (CCSSO), the Science Assessment Item Collaborative (SAIC) was formed in 2015, and in collaboration with WestEd, produced guiding documentation for use by both state members and the public for Next Generation Science Standards (NGSS) large-scale assessment production in the form of an Assessment Framework ("Assessment Framework"), Item Specifications Guidelines, and Item Cluster Prototypes. The Assessment Framework provides a range of options and accompanying rationales for the development of NGSS-aligned items and large-scale summative assessments, and was designed to be used in concert with the Item Specifications Guidelines to aid state education agencies (SEAs) and other entities in documenting the processes needed to drive the development of NGSS-aligned items and assessments. These documents and prototypes address the major issues facing SEAs and other entities implementing new science standards by documenting the processes needed to guide the development of large-scale summative assessments for the NGSS. Since the date of their publication, these documents have been leveraged by states to guide the development of their NGSS-aligned state summative assessments. In doing so, much has been learned about the processes and principles that were originally outlined in the Assessment Framework and Item Specifications Guidelines. This report provides an update the assessment community on the insights and key takeaways learned during initial implementation of these SAIC resources.

INTRODUCTION

A year after the CCSSO Science Assessment Item Collaborative (SAIC) Assessment Framework was released, WestEd was given the opportunity to survey original SAIC members on how states have used SAIC materials or engaged in other activities related to NGSS large-scale assessment. To this end, WestEd administered a survey in December, 2016 designed to determine:

- how states have used the original SAIC materials and how those uses have diverged from the original documentation;
- information for drafting an addendum to the original documentation, based on the practical experiences of states.

The survey was distributed to individual SAIC members. A total of eight SAIC states responded to WestEd's survey.

The first section of this addendum details specific aspects of the SAIC Assessment Framework that merit adjustment in scope or content based on member states' feedback.

The second section of this addendum aggregates the results and findings from the survey.

SUGGESTED ADJUSTMENTS TO SAIC ASSESSMENT FRAMEWORK RECOMMENDATIONS

The primary focus of the Assessment Framework was to build a basis of item development for NGSS large-scale assessment within the context of overall test design. The Assessment Framework was intended to be considered a starting point for the implementation of a large-scale assessment measuring the NGSS, rather than being considered the final model. The Assessment Framework noted that the item cluster model presented in the Assessment Framework had not been developed and fully implemented in a state testing system for science at the time of publication. The Assessment Framework acknowledged that lessons learned through large-scale development would present opportunities to adjust the model presented and tools recommended. The descriptions and expectations presented in the Assessment Framework were to be considered a starting point, rather than the definitive end product. The following table identifies and details specific aspects of the SAIC Assessment Framework and Item Specifications Guidelines that merit adjustment in scope or content based on member state feedback.

Reference	Text from Document	Proposed Update
Assessment Framework: Chapter 4, p.22	Item clusters, not individual items, are the base unit for the SAIC test development. That is, individual items are intentionally developed to be situated within the context of an item cluster and not to be used as stand-alone items.	Item clusters are still considered to be the base unit for SAIC test development, with three-dimensional alignment being achieved across a full item cluster. However, consideration should be paid to individual items being used as standalone, or discrete items in overall test construction. This allows for some flexibility in test development, to achieve a wider range of dimensions and PEs, than is afforded with a test built entirely of item clusters.

Reference	Text from Document	Proposed Update
Assessment Framework: Chapter 4, p.23	Each item is inextricably linked to the stimulus and to the other items within the item cluster, and the stimulus may be interspersed among the items to add information as needed. This means that student exposure to the stimulus is considered essential in order to respond correctly to any individual item, and that the item cluster must be constructed in such a way that individual performance on each item is adversely affected if an item is responded to without the context of the other items in the cluster.	Consideration should be paid to the stringency pertaining to an item being adversely affected without the context of other items in the cluster. Although this represents one approach that may be favored in some circumstances, an approach that provides greater flexibility to lose or substitute items may be considered for pragmatic reasons.
Assessment Framework: Chapter 4, p.27	Hot Spot (Sample Task/Purpose): Identify aspects of a model that support a given claim	Add "Build/develop a model" to Sample Task column.
Assessment Framework: Chapter 4, p.25	While it may be possible to develop items within a single cluster that are collectively sufficient to assess the entirety of a single PE, this is not preferable and will not be possible in many, if not most, cases. For item clusters inclusive of more than one PE, it is not expected that a single item cluster will be able to provide the opportunity for a student to generate evidence of every aspect of each PE in the item PE bundle.	Add consideration for the possibility of bringing in SEPs that are not specifically linked to the bundled (or individual) PE. Also may note in this section that the possibility of using CCCs as "anchors" for PE bundles (and therefore as themes for the item clusters) is a feasible approach.

Reference	Text from Document	Proposed Update
Assessment Framework: Chapter 7, p.44	The stimulus is an integral part of the item cluster and, therefore, should meet specific criteria for inclusion in an NGSA. To mirror the work of real scientists, the stimulus should be based on a real-world science phenomenon that is representative of how students learn in the classroom. The ideal stimulus also should have a rich, grade-appropriate context that can support a variety of robust item types used to gather evidence of what students know and are able to do.	Confining an item cluster stimulus to a single natural phenomenon may be too restrictive. The stimulus could involve a broader scenario (e.g., presenting details of how an ecosystem has changed over time) that will encompass multiple phenomena that can be examined via individual items within the cluster.
Item Specifications Guidelines: Appendix A, p.35	The stimulus for an item cluster must be broad enough in content to support all of the items in the cluster yet flexible enough for students to exhibit their ability to demonstrate their capabilities to apply SEPs. Because many item clusters may require students to demonstrate capacity to develop and use models, plan and carry out an investigation, and subsequently interpret the data and construct explanations based on the data, the stimulus may need to include information that is extraneous or tangential to the overall goal in order for students to demonstrate their capacity to identify appropriate or pertinent information or data from a stimulus.	Detail modeling approaches: searching for appropriate models can be a productive starting point. The search for models can begin with scientific literature, and the developer can evaluate potential models using the lens of whether the model can be simplified appropriately for grade-level (amount of information to digest) and PE boundaries while still maintaining scientific relevance.

Reference	Text from Document	Proposed Update
Item Specifications Guidelines: Appendix A, p.35	The stimulus for an item cluster must be broad enough in content to support all of the items in the cluster yet flexible enough for students to exhibit their ability to demonstrate their capabilities to apply SEPs. Because many item clusters may require students to demonstrate capacity to develop and use models, plan and carry out an investigation, and subsequently interpret the data and construct explanations based on the data, the stimulus may need to include information that is extraneous or tangential to the overall goal in order for students to demonstrate their capacity to identify appropriate or pertinent information or data from a stimulus.	Emphasis should be placed on the critical importance of including data to standalone item development, and also about potential uses of simplified scientific models as a starting point for standalone item development, especially for the modeling SEP and systems CCC. The importance of data searches in deciding on a stimulus context—a context that may seem to work well in theory or in a general sense is not sufficient if not supported by appropriate data—can lead to "dead ends" in stimulus development work. Searching for contexts that are supported by multiple types and sources of data that are relevant and accessible to students, as well as within all constraints and boundaries of the PE greatly facilitates the development of the IC by allowing for flexibility in approaches that can support different combinations of dimensions and practices and will ultimately result in the desired range of item alignments.

SURVEY QUESTIONS

The following questions were included in the SAIC survey:

1. *Which description best fits your state's NGSS adoption model?*

2. *To what degree has your state used the Assessment Framework (including the Item Specifications Guidelines) since its release?*

3. *Describe how your state has used the Assessment Framework since its release.*

4. *Has your state used aspects of the Assessment Framework and/or Prototypes in RFPs?*

5. *Has your state used aspects of the Assessment Framework and/or Prototypes in directing the work of your current development vendor(s)?*

6. *Describe any modifications your state has made to the Assessment Framework (or underpinning principles) for use in your state.*

7. *To what degree has your state used the Item Cluster prototypes since their release?*

8. *Describe how your state has used the Item Cluster prototypes since their release.*

9. *Describe any modifications your state has made to the Item Cluster model (or underpinning principles) for use in your state.*

10. *To what degree has your state used the NGSS Evidence Statements in your assessment development work?*

11. *Describe how your state has used the NGSS Evidence Statements.*

12. *Which of the following item types has your state developed, or intends to develop, for your state's large-scale science assessment? Select all that apply.*

13. *When does your state anticipate administering its first operational NGSS assessment?*

14. *How far has your state's assessment development progressed since new standards were adopted?*

15. *Describe the item clusters your state has developed (if applicable).*

16. *How closely do you feel the item cluster model used in your state matches the model and prototypes presented as part of the SAIC Phase 1.0 work?*

17. *Has your state field tested NGSS-aligned Item Clusters?*

18. *Has your state conducted cognitive laboratory studies with item clusters?*

19. *Would your state be willing to share that data and results with other states?*

20. *What resources have your state developed/used in developing NGSS assessments? Which of these resources would your state be willing to share with other states?*

21. *Has your state developed any of the following for NGSS-aligned assessments?*
    ☐ *Blueprints*   ☐ *Alternate Assessments*   ☐ *Parent/Student Reports*
    ☐ *Claims*   ☐ *Achievement Level Descriptors*

22. *Is your state in the process of developing any of the following for NGSS-aligned assessments?*
    ☐ *Blueprints*   ☐ *Alternate Assessments*   ☐ *Parent/Student Reports*
    ☐ *Claims*   ☐ *Achievement Level Descriptors*

23. *Describe how, or to what degree, your state has implemented a comprehensive science assessment system.*

24. *Describe some of the challenges your state has faced in implementing an NGSS large-scale assessment and the paths taken to overcome those challenges.*

25. *Describe input/questions/concerns from your Technical Advisory Committee related to an NGSS large-scale assessment that other states could benefit from.*

26. *What are the most significant cost drivers impacting the achievement of your envisioned NGSS large-scale assessment (e.g., constructed response scoring, development, delivery systems)?*

27. *What, if any, collaborations among other states for NGSS large-scale assessment has your state been involved in?*

28. *What workaround accommodations and accessibility specific to NGSS large-scale assessments has your state engaged in?*

29. *What do you see as the most critical next steps facing the field relative to NGSS large-scale assessment?*

30. *Is there any additional information you would like to share regarding use of the SAIC materials or about NGSS large-scale assessment?*

RESPONSES FROM SURVEY PARTICIPANTS

1. *"Which description best fits your state's NGSS adoption model?"*

   In terms of NGSS adoption model, 50 percent of the responding SAIC member states have fully adopted the NGSS, with a name change. 38 percent of the responding states fully adopted NGSS by name, while 12 percent adapted NGSS to meet state-specific needs.

   Responses:
   - Full adoption of NGSS by name: 3 – 38%
   - Full adoption with name change: 4 – 50%
   - Adaptation (adapted NGSS to state specific needs): 1 – 12%
   - Adaptation (update existing standards under principles of NGSS): 0 – 0%

2. *"To what degree has your state used the Assessment Framework (including the Item Specifications Guidelines) since its release?"*

   In terms of how states have used the Assessment Framework, 50 percent of the responding SAIC member states have used the materials as a starting point, and subsequently built on the materials. 38 percent of responding states have used the materials to a minimal degree, and 12 percent have used the materials extensively in their work.

   Responses:
   - Our state has not used any of the materials: 0 – 0%
   - Our state has used the materials to a minimal degree (e.g., reviewed internally, discussed internally): 3 – 38%
   - Our state has used the materials as a starting point and has built on them: 4 – 50%
   - Our state has used the materials extensively in our work (e.g., in training, in ongoing development efforts): 1 – 12%

3. *"Describe how your state has used the Assessment Framework since its release."*
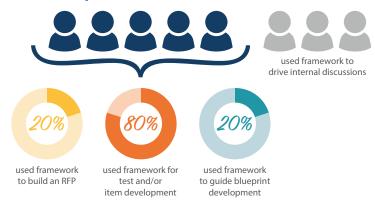
Since its release, some responding SAIC member states have used the Assessment Framework as part of discussions around assessment development, with one state using the framework to guide development of state NGSS test blueprints. Other states have used the Assessment Framework in the development of item clusters. Another state has used the Assessment Framework in building their social studies RFP.

Responses:

- One SAIC member used information from the Assessment Framework to guide thinking on the development of state NGSS test blueprints. This member shared the Framework with its Technical Advisory Committee as an introduction into test blueprint conversations.

- One SAIC member has used the Assessment Framework for test development.

- One SAIC member has used the Assessment Framework in internal discussions and as a guide to help inform work moving forward with NGSS assessment development.

- One SAIC member has incorporated some of the ideas in the framework, such as clusters of items (referred to by the member as "modules").

- One SAIC member has used the item cluster ideas and some of the item specifications with respect to 2D and 3D items.

- One SAIC member has used the framework in building their first science and social studies

Figure 1: **How responding SAIC member states are using the Assessment Framework**

5 of 8 states used the Assessment Framework (including the Item Specifications Guidelines **extensively** or as a **starting point to then build upon**.

used framework to drive internal discussions

20% used framework to build an RFP

80% used framework for test and/or item development

20% used framework to guide blueprint development

RFP. The state then rethought its approach and developed a different science model, but is using the SAIC approach for its social studies assessment.

- One SAIC member is following some of the general guidance, including the development of item clusters.

4. *"Has your state used aspects of the Assessment Framework and/or Prototypes in RFPs?"*

Most of the responding SAIC member states (62 percent) have not used aspects of the Assessment Framework or Prototypes in their RFPs. The other 38 percent have used aspects of the Assessment Framework or Prototypes in their RFPs.

Responses:

- Yes: 3 – 38%
- No: 5 – 62%

5. *"Has your state used aspects of the Assessment Framework and/or Prototypes in directing the work of your current development vendor(s)?"*

Many of the responding SAIC member states (75 percent) have used aspects of the Assessment Framework and/or Prototypes in directing the work of their current development vendor. 25 percent of responding states have not used aspects of the Assessment Framework and/or Prototypes in directing the work of their current development vendor.

Responses:
- Yes: 6 – 75%
- No: 2 – 25%

6. *"Describe any modifications your state has made to the Assessment Framework (or underpinning principles) for use in your state."*

With regards to modifications made to the Assessment Framework, one responding SAIC member found that there should be a close relationship between a stimulus and other items within a cluster, but that there should be the ability to "lose" an item without losing the entire cluster.

Responses:
- One SAIC member found one of the claims (*Bundled PEs should have at least one dimension in common*) to be "not necessarily true." This state also found that there should be a close relationship between a stimulus and other items within a cluster, but that there should be the ability to "lose" an item, based on a review of pilot data, and not "lose" the entire cluster.
- One SAIC member is implementing a 2D model, without crosscutting concepts.
- One SAIC member has used the prototypes as examples of what an item cluster can look like.
- One SAIC member is not using the SAIC prototypes for science; while the state might use them, it will not be used as the primary approach. However, the state is using this model for its social studies assessment to incorporate Common Core standards with social studies standards, making items/clusters 2D in nature.
- One SAIC member is using the assessment framework as one resource to develop state science assessments.

7. *"To what degree has your state used the Item Cluster prototypes since their release?"*

Most of the responding SAIC member states (63 percent) have used the Item Cluster prototypes as a starting point and built on them. 25 percent of responding states have used the prototypes to a minimal degree. 12 of the responding states have used the prototypes extensively.

Responses:

- Our state has not used the prototypes: 0 – 0%

- Our state has used the prototypes to a minimal degree (e.g., reviewed internally, discussed internally): 2 – 25%

- Our state has used the prototypes as a starting point and has built on them: 5 – 63%

- Our state has used the prototypes extensively in our work (e.g., in training, in ongoing development efforts): 1 – 12%

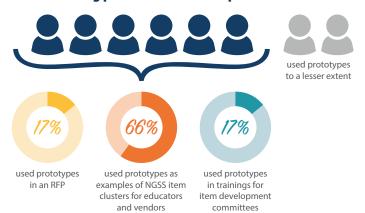**8.** *"Describe how your state has used the Item Cluster prototypes since their release."*

Some responding SAIC members have used the Item Cluster prototypes as illustrative examples of NGSS-aligned clusters, with some sharing them with their item development committees and their assessment vendor. One state used the Item Cluster prototypes in an RFP.

Responses:

- One SAIC member has made the prototypes part of initial training for item development committees, especially early on when very few of the committee members had seen them. For this state, the prototypes provided a good

Figure 2: How responding SAIC member states are using the Item Cluster prototypes

6 of 8 states used the Item Cluster prototypes **extensively** or as a **starting point to then build upon**.

used prototypes to a lesser extent

17% used prototypes in an RFP

66% used prototypes as examples of NGSS item clusters for educators and vendors

17% used prototypes in trainings for item development committees

starting place for discussion and for thinking about how assessment can/should be different for the NGSS.

- One SAIC member has used the Item Cluster prototypes with educators and vendors as an example of what NGSS clusters might look like.

- One SAIC member reviewed the Item Cluster prototypes internally, as well as shared them with their vendor for ideas on how to combine the practices and content of the standards.

- One SAIC member has used them as examples.

- One SAIC member has used the prototypes in a cancelled RFP, as well as in a social studies RFP as examples.

- One SAIC member has used the grade 5 prototype as an example of an NGSS-aligned item cluster with various audiences.

9. *"Describe any modifications your state has made to the Item Cluster model (or underpinning principles) for use in your state."*

One responding SAIC member modified the Item Cluster models by making it so that a 2D or 3D item can be aligned to a combination of dimensions of PEs within a bundle (e.g., if PE1 and PE2 are bundled, a 2D item can be aligned to the SEP from PE1 and the DCI from PE2). Another state is considering using shorter clusters around a single PE, rather than bundling.

Responses:
- One SAIC member has made it so that a 2D or 3D item can be aligned to a combination of dimensions of PEs within a bundle (e.g., if PE1 and PE2 are bundled, a 2D item can be aligned to the SEP from PE1 and the DCI from PE2).
- One SAIC member is considering shorter clusters around a single PE rather than bundling.
- One SAIC member is not using 3D item clusters.
- One SAIC member has not worked with their social studies team beyond writing the RFP – further development has just begun.
- One SAIC member is developing a variety of item cluster types, including ones that are shorter and aligned to only one Performance Expectation.

10. *"To what degree has your state used the NGSS Evidence Statements in your assessment development work?"*

50 percent of the responding SAIC member states have used the NGSS Evidence Statements (NGSS Network, 2015) to a minimal degree in their assessment development work. 38 percent of responding states have used the NGSS Evidence Statements extensively in their assessment development work, and 12 percent have used them as a starting point, with adaptation.

Responses:
- Not at all: 0 – 0%
- To a minimal degree: 4 – 50%
- As a starting point, with adaptation: 1 – 12%
- Used extensively: 3 – 38%

11. *"Describe how your state has used the NGSS Evidence Statements."*

The NGSS Evidence Statements have been used as guidance for item specifications by some of the responding SAIC members. The NGSS Evidence Statements have also been used in professional development. Two states have used the NGSS Evidence Statements as a reference. One state hopes to only use the Evidence Statements as clarifiers for item writers looking for assistance, and not for item development or instructional guidance.
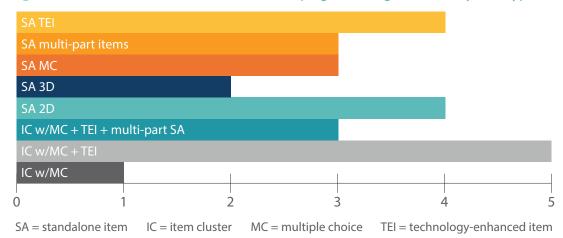
Responses:

- In the first rounds of item development, one SAIC member used the evidence statements essentially as item specifications. Since the development of their item specifications is not yet complete, this state is still depending on the evidence statements for guidance, but taking note of where and when item specifications differ.

- One SAIC member uses the NGSS Evidence Statements extensively in professional development, and will do the same with item development.

- One SAIC member has looked at the evidence statements internally in hopes of using them for future development. Currently, this state is using resources from a partner state and has not completed any independent development.

- One SAIC member looks to verify with the NGSS Evidence Statements on occasion, but this state has different standards so the two do not always match up.

- One SAIC member has only used the NGSS Evidence Statements as a reference.

- One SAIC member hopes to only use the Evidence Statements as clarifiers for item writers looking for assistance in writing items. The Evidence Statements are not to be used as primary sources of information or guides for item development, and are not used in the classroom to specifically guide instruction.

- One SAIC member has aligned items to the evidence statements and also used them as the basis to develop task demands in item specifications that describe the types of questions to be developed.

12. *"Which of the following item types has your state developed, or intends to develop, for your state's large-scale science assessment? Select all that apply."*

Responding SAIC members indicated that the following item types have been developed, or will be developed, for a state's large-scale assessment:

Figure 3: **Number of SAIC member states developing NGSS-aligned items, by item type**

SA = standalone item   IC = item cluster   MC = multiple choice   TEI = technology-enhanced item

13. *"When does your state anticipate administering its first operational NGSS assessment?"*

Responding SAIC member states indicated that their NGSS assessments are planned to be operational sometime between 2017–2020. One state is planning to administer their NGSS assessment in 2017, one in 2018, three in 2019, and two in 2020.

Responses:

- Pilot assessment (feasibility) work in 2016–17, field testing in 2017–18, and live assessment in 2018–19
- 2017
- Spring 2018
- 2019
- 2019
- 2020
- Spring 2020

14. *"How far has your state's assessment development progressed since new standards were adopted?"*

In terms of states' assessment development progress, 43 percent of responding SAIC member states have progressed to prototypes, while another 43 percent have progressed to field testing. 14 percent of responding states have not progressed to prototypes, field testing, or operational testing.

Responses:

- Prototypes: 3 – 43%
- Field testing: 3 – 43%
- Operational testing (producing scores as reported to students): 0 – 0%
- None of the above: 1 – 14%

Figure 4: **Percentage of responding states anticipating the administration of NGSS-aligned tests, by year**

3 states have progressed to **field testing** using the new standards

15. *"Describe the item clusters your state has developed (if applicable)."*

The responding SAIC members are in various stages of item cluster development. Two of the states have created clusters, with one state basing a quarter of their clusters around a single PE, and two clusters around a 3-PE bundle. One state may move to single PE short clusters. Three states are currently developing item clusters, with one planning to release grades 5 and 8 prototypes in the fall for practice.

Responses:

- One SAIC member has created clusters most often based on a bundle of two PEs. Approximately one-quarter of the clusters are based on a single PE. This state has two item clusters that are based on a three-PE bundle. The clusters have ranged from three to six items each, with more than half of those items being multi-part. As this state has moved from a paper-based format to an online format, it has discovered that the variety of item types and the ability to write multi-part items have allowed for far less dependence on short answer items to gather evidence of high level thinking.

- One SAIC member has a limited number of item clusters developed at this time. The state may move to single PE short clusters.

- One SAIC member is currently developing prototypes, which will be reviewed with teachers in January, tried out in June, and

released in the fall for practice purposes for grades 5 and 8. Development for high school is a year behind.

- One SAIC member is developing item clusters that assess topic bundles of PEs. Each cluster contains between five and eight questions.

- One SAIC member is developing a variety of item types. Some are short and aligned to only one PE, others are longer and aligned to two PEs.

16. *"How closely do you feel the item cluster model used in your state matches the model and prototypes presented as part of the SAIC Phase 1.0 work?"*

Most of the responding SAIC members found somewhat of a match between their state item cluster model and the SAIC Phase 1.0 model and prototypes. One state found a close match between the two.

Responses:

- One SAIC member finds the state's item cluster model to match closely to the SAIC model and prototypes.

- One SAIC member finds the state's item cluster model similar, but not exact with the SAIC model and prototypes. This state includes constructed, hand-scored three point responses.

- Three SAIC members see somewhat of a match between their item cluster model and that of the SAIC model and prototypes.

17. *"Has your state field tested NGSS-aligned Item Clusters?"*

The majority of responding SAIC member states (71 percent) have not field tested NGSS-aligned Item Clusters. 29 percent of responding states have field tested NGSS-aligned Item Clusters.

One state that field tested NGSS-aligned Item Clusters found that the items were quite difficult for students, compared to the operational items based on previous standards. This state found that their students struggled to use appropriate evidence to support claims and describe limitations of models and relationships between model components. Students also had trouble stating and evaluating claims based on evidence from one or more sources.

One state will field test NGSS-aligned Item Clusters in the spring of 2017.

Responses:
- Yes: 2 – 29%
- No: 5 – 71%

High-level takeaways from data analysis:
- One SAIC member found that the items were quite difficult for students, compared to the operational items based on previous standards. Students struggled to use appropriate evidence to support claims and to describe limitations of models and relationships between components of models. Students also had trouble stating and evaluating claims based on evidence from one or more sources. Far fewer short answer items were piloted, compared to usual, due to new more complex items types being useful in gathering evidence of higher level thinking.

18. *"Has your state conducted cognitive laboratory studies with item clusters?"*

None of the responding SAIC member states have conducted cognitive laboratory studies with item clusters. Cognitive laboratory studies are in planning stages for three states, with one planning to conduct them in the spring of 2017.

Responses:
- Yes: 0 – 0%
- No: 7 – 100%

Open-ended responses:
- For one SAIC member, cognitive laboratory studies are in the planning stages.
- One SAIC member will conduct cognitive laboratory studies in spring of 2017.
- One SAIC member may conduct cognitive laboratory studies this year as part of pilot work for a large scale system of state assessments.

19. Would your state be willing to share that data and results with other states?

(No responses)

20. *"What resources have your state developed/used in developing NGSS assessments? Which of these resources would your state be willing to share with other states?"*

Of the responding SAIC member states, the following resources have been used in developing NGSS assessments: Item Writers' Guide; item development training presentations for item writing, content review, content review with data, and pilot range finding meetings; a brief handout detailing the "anatomy of a standard" and the "anatomy of an evidence statement"; alignment documents for making cluster and standalone alignment judgments; unpacking documents; teacher training tools; and item templates. For two of the responding SAIC member states, resources are currently in draft form. Two of the members indicated a willingness to share resources with other states.

Responses:

- One SAIC member has used the following resources: Item Writers' Guide; item development training presentations for item writing, content review, content review with data, and pilot range finding meetings; a brief handout detailing the "anatomy of a standard"

and the "anatomy of an evidence statement;" and alignment documents for making cluster and standalone alignment judgments.

- For one SAIC member, most resources are still in draft or infancy form.

- One SAIC member would be happy to share once they have received the data on tryout and release items, expected sometime in fall of 2017.

- One SAIC member has used: unpacking documents, teacher training tools, item templates, etc. This member would be willing to share their resources with other states.

- One SAIC member is closely following the advice of the BOTA report (NRC, 2014), which describes the development of a multi-layered system of assessments that provide feedback to teachers and students throughout the instructional setting, while also providing an audit of science literacy at pivotal points in K–12 education (for this state, grades 5, 8, and 10).

- One SAIC member is in the process of developing item specifications and documents related to test design and reporting. At this point, the resources are still in draft form and secure.

21. *"Has your state developed any of the following for NGSS-aligned assessments?"*

In terms of items developed for NGSS-aligned assessments, three states have developed blueprints, and claims have been developed in three states.

22. *"Is your state in the process of developing any of the following for NGSS-aligned assessments?"*

In terms of documentation currently being developed for NGSS-aligned assessments, four states are developing blueprints and three are developing claims. Parent/student reports are being developed in two states. Achievement Level Descriptors are being developed in three states. Four states are in the process of developing alternate assessments.

Figure 5: **Number of SAIC responding states developing NGSS-aligned materials**

■ **currently developed** NGSS-aligned material
■ NGSS-aligned material **under construction**

23. *"Describe how, or to what degree, your state has implemented a comprehensive science assessment system."*

None of the responding SAIC member states have implemented a comprehensive science assessment system, though half of the respondents indicated plans to do so. Another state has started development of classroom embedded assessments, working with its vendor to assess the feasibility of items for end-of-unit assessments and integrative transfer task audit assessments for classroom and school/state use. Two of the respondents indicated budgetary constraints that have impacted development of a comprehensive assessment system.

Responses:

• One SAIC member is hoping for legislative funding to support work on the comprehensive science assessment system. As item development allows, this state will make some item clusters available for district use in non-tested grades.

• One SAIC member is hoping to have a comprehensive system developed collaboratively by its Assessment staff and the Curriculum and Instruction group.

• One SAIC member tests in grades 5, 8, and high school (grade 9 or 10); administers EOC high school tests in Biology, Chemistry, Intro Physics, and Tech/Eng; and is currently modifying this selection for the next generation science tests.

• One SAIC member is working on its development.

• One SAIC member has begun development of classroom embedded assessments created

by teachers, for teachers. Work has been done with the assistance of consultants. The state is now working with its vendor to assess the feasibility of items for end-of-unit assessments and integrative transfer task audit assessments for classroom and school/state use.

- One SAIC member is focusing current efforts on state summative assessment development. The state hopes to make additional resources for formative and interim assessments available, which may be developed by a variety of sources. Specific development work on these resources has been stalled due to budget constraints.

24. *"Describe some of the challenges your state has faced in implementing an NGSS large-scale assessment and the paths taken to overcome those challenges."*

The challenges of implementing a large-scale NGSS assessment range from development issues to transition issues to public response issues. Developing aligned items has been a "steep learning curve" due to the complexity of the standards, as well as working within federal rules regarding assessment. Communication around a new assessment has also been a challenge due to public resistance to large-scale assessment. To address the challenges associated with working with new standards, one state has created and released an assessment "crosswalk" that highlights overlapping standards.

Responses:

- One SAIC member found that challenges primarily resulted from the complexity of the standards and the time required for all to develop a deep understanding of what those standards look like in both good science instruction and good assessment. This state found there is a steep learning curve for everyone involved, from the teachers on the item development committee to the content specialists with many of the vendors.

- One SAIC member found challenges in terms of transitioning to new standards while still assessing "old" standards, and grade span testing versus grade level testing. This member released a document with the tested standards for 2017 in grades 5 and 8, and will also release an assessment "crosswalk" (planned for release in early 2017) of overlapping standards for testing purposes that will be used for the 2018 test plan.

- For one SAIC member, this is ongoing. Their challenges include ongoing resistance to large-scale assessment, and a need for ongoing communication and education of public and internal stakeholders, all while working with psychometricians and policymakers to navigate what can and cannot be done within the "rules" of the federal assessment system.

25. *"Describe input/questions/concerns from your Technical Advisory Committee related to an NGSS large-scale assessment that other states could benefit from."*

Based on conversations with their Technical Advisory Committees (TAC), responding SAIC members have identified a number of concerns regarding NGSS large-scale assessment. In one state, these concerns centered on the specificity of information that can be derived from data, the feasibility of item independence, and the pros and cons of matrix sampling. In another state, the TAC raised concerns about test length and accessibility. Two states have had positive conversations with their TAC regarding their plans, though one state acknowledged the possibility of federal pushback due to the state's nontraditional assessment design.

Responses:
- The TAC from one SAIC member has psychometric-based concerns. The multi-dimensionality of the standards and the idea of item clusters/PE bundles have raised questions about: the specificity of information that can be derived from the data, the feasibility of item independence, and the pros and cons of matrix sampling.
- One SAIC member lists concerns regarding test length and accessibility.
- One SAIC member has not reached this point yet, and hopes to do so in the spring.
- The TAC from one SAIC member is very supportive of the state's plan.
- One SAIC member is conducting all work under advisement of its TAC. To date, all

comments have been positive, but with acknowledgement that there may be some pushback to the nontraditional nature of the state's design.

26. *"What are the most significant cost drivers impacting the achievement of your envisioned NGSS large-scale assessment (e.g., constructed response scoring, development, delivery systems)?"*

In terms of significant cost drivers that have impacted states' envisioned NGSS large-scale assessment, responding SAIC members identified a variety of factors: scoring, development costs, delivery systems, and holding teacher meetings.

Responses:
- One SAIC member thinks that the largest cost driver will be around development costs. This state is only able to get about half the number of items from an item development committee than it was able to get in the past. Developing stimuli and items for an item cluster is complex and very time-consuming.
- One SAIC member identified scoring and holding teacher meetings.
- One SAIC member identified CR scoring.
- One SAIC member identified scoring methodologies and online scenario development, with requirement to have constant replacements due to item memorability.
- One SAIC member identified constructed response scoring, development, and delivery systems.

27. *"What, if any, collaborations among other states for NGSS large-scale assessment has your state been involved in?"*

Collaboration has occurred among responding SAIC member states around NGSS large-scale assessment. Some of this collaboration has occurred among NGSS states, including some collaboration among states that share the same assessment vendor. One state has found itself with fewer direct collaborators after the state elected to develop a full-scale assessment system, with no other state doing the same.

Responses:
- One SAIC member has been engaged in regular (twice a month) conversations with several other NGSS-adopter states that share the same vendor. This state has been able to review a small number of items from these other states but is working to facilitate this process through the vendor's item banking system. This state is investigating item sharing agreement possibilities.

- One SAIC member will be part of one or more emerging collaboratives.

- One SAIC member has had some collaboration with other states.

- One SAIC member was collaborating with everyone, but once this state decided to go down the path of developing a full-scale assessment system, there were no other states doing the same. This has left the state with fewer direct collaborators. This does not mean the state is not willing to share what it has done, or that it is unwilling to learn from others to incorporate what they are doing. The state is especially interested in the work

being done in another state, as their formative assessment work is similar to the embedded assessment work this state is doing. At some point, this state would be interested in trying to develop an online library of NGSS-aligned items that teachers could use freely within their classrooms to assess their students appropriately. This would require vetting of items, so a team of highly trained teachers or assessment leads would be needed for that job.

- One SAIC member is collaborating with several other states to share NGSS assessment resources and items using a common testing contractor.

28. *"What work around accommodations and accessibility specific to NGSS large-scale assessments has your state engaged in?"*

In terms of workaround accommodations and accessibility for NGSS large-scale assessments, two responding SAIC member states indicated their plans. One state will use the most accessible item types, while the other state will try to match Smarter Balanced Mathematics supports and accommodations, though supports for ASL may not be possible due to cost. Another state has started to work with their alternate assessment group.

Responses:
- One SAIC member has endeavored to use the most accessible item types possible for each item.

- One SAIC member will try to match Smarter Balanced Math supports and accommodation, although ASL may or may not be possible due to cost.

- One SAIC member has started to work with their alternate assessment group.

29. *"What do you see as the most critical next steps facing the field relative to NGSS large-scale assessment?"*

For the most critical next steps relative to NGSS large-scale assessment, the responding SAIC members had a wide range of responses. Some of the next steps identified concerns about test development: developing test blueprints and high-quality items, increasing access to practice tests, as well as making a reasonable validity argument for the use of item clusters. One state would also like to see a common vision and excellent coordination among states that are interested in sharing items. Public relations/communication was also listed as a consideration, as states will need to communicate the value of assessments as part of improving teaching and learning, particularly the important role of classroom assessments. One state is concerned about federal assessment requirements; in particular, this state would like for its field test to count one year for its accountability system to avoid administering two science assessments.

Responses:
- One SAIC member sees the need for a common vision and excellent coordination among states that would like to share items.

- One SAIC member sees public relations around the value and utility of the assessment and the contribution of an assessment system to overall performance and the enhancement of both teaching and learning as a critical next step.

- One SAIC member sees the need to highlight: the importance of classroom assessments, which need improvement too, as large-scale testing; and access to practice tests, especially the online tools that students are becoming familiar with.

- One SAIC member sees the need to make a reasonable validity argument for the use of item clusters.

- One SAIC member identifies many critical next steps. One that the state would like to see occur is for the feds to allow the state's field test to count as "the test" for one year so that the state does not have to give double science assessments (creating public pushback for no good reason) in 2017–18. Another step that would be helpful is to get a team of open minded psychometricians willing to help the state brainstorm through the roadblocks to come up with nontraditional evaluations.

- One SAIC member identified developing test blueprints and enough high quality items to support the design.

30. *"Is there any additional information you would like to share regarding use of the SAIC materials or about NGSS large-scale assessment?"*

Two of the responding SAIC members identified participation in SAIC as an important starting point for their work in developing a large-scale NGSS assessment, sparking important discussions and initial work for assessment development.

Responses:

- One SAIC member shared that participation in SAIC gave their state a starting point for getting the development of large scale NGSS assessment off the ground. The Assessment Framework provided the structure needed to begin writing item clusters within their team and with teacher committees, and helped start important discussions around test design with its TAC.

- One SAIC member noted that without SAIC as an evolutionary step in thinking, the state would not have come to the place that they now find themselves. It took that position between traditional and a bit far out to help their administration take the leap to what the BOTA report actually suggested. The SAIC has been an incredibly important process and whether teams use the outcomes of that process, or if they jump off from that point to something even further afield, none of it would have happened without the SAIC process.

REFERENCES

CCSSO and WestEd. 2015. *Science Assessment Item Collaborative Assessment Framework for the Next Generation Science Standards*. Washington, DC: The Council of Chief State School Officers.

CCSSO and WestEd. 2015. *Science Assessment Item Collaborative Item Specifications Guidelines for the Next Generation Science Standards*. Washington, DC: The Council of Chief State School Officers.

CCSSO and WestEd. 2015. *Grade 5 Item Cluster Prototype*. Washington, DC: The Council of Chief State School Officers.

CCSSO and WestEd. 2015. *High School Item Cluster Prototype*. Washington, DC: The Council of Chief State School Officers.

National Research Council (NRC). 2014. *Developing assessments for the Next Generation Science Standards. Committee on Developing Assessments of Science Proficiency in K–12*. Board on Testing and Assessment and Board on Science Education, J. W. Pellegrino, M. R. Wilson, J. A. Koenig, & A. S. Beatty (Eds.). Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Next Generation Science Standards (NGSS) Network. 2015. *NGSS Evidence Statements*. Retrieved from http://www.nextgenscience.org/evidence-statements

NGSS Lead States. 2013. *Next Generation Science Standards: For States, By States.* Washington, DC: The National Academies Press.

CSAI provides state education agencies (SEAs) and Regional Comprehensive Centers (RCCs) with research support, technical assistance, tools, and other resources to help inform decisions about standards, assessment, and accountability. CSAI is committed to: 1) Improving state policies, practices, and supports to districts for the implementation of rigorous college and career readiness standards and assessments; 2) Increasing the capacity of the RCCs to meet the technical assistance and implementation needs of states around standards, assessment, and accountability; 3) Capitalizing on available technology and media to promote engagement, effectiveness, and efficiency in the application of research and resources to expand the reach of CSAI; and 4) Continuously improving the operational efficiency and effectiveness of CSAI to support the long-term sustainability of educational reform efforts across states and districts.

WestEd is a nonpartisan, nonprofit research, development, and service agency that partners with education and other communities throughout the United States and abroad to promote excellence, achieve equity, and improve learning for children, youth, and adults. WestEd has more than a dozen offices nationwide, from Massachusetts, Vermont and Georgia, to Illinois, Arizona and California, with headquarters in San Francisco.

For more information, visit WestEd.org; call 415.565.3000 or, toll-free, (877) 4-WestEd; or write: WestEd / 730 Harrison Street / San Francisco, CA 94107-1242.