Alignment Considerations for Next Generation Science Standards Assessments

September 2017

THE CENTER ON
STANDARDS & ASSESSMENT IMPLEMENTATION
WestEd CRESST

ACKNOWLEDGEMENTS

We gratefully acknowledge the Council of Chief State School Officers (CCSSO) for their role and leadership in the Science Assessment Item Collaborative (SAIC). We wish to thank the following members of the WestEd science assessment team for their expertise: Erika Gasper, Kevin King, Thomas Parham, Cinda Parton, and Nicolle Romero.

THE CENTER ON STANDARDS & ASSESSMENT IMPLEMENTATION
WestEd CRESST

WestEd
WestEd.org

The work reported herein was supported by grant number #S283B050022A between the U.S. Department of Education and WestEd with a subcontract to the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). The findings and opinions expressed in this publication are those of the authors and do not necessarily reflect the positions or policies of CRESST, WestEd, or the U.S. Department of Education.

This material is based upon work supported by the CCSSO SAIC. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the SAIC or CCSSO.

ABSTRACT

Many states have already undertaken efforts to develop large-scale assessments aligned to the Next Generation Science Standards (NGSS) (NGSS Lead States, 2013) and other science standards inspired by A Framework for K-12 Science. More than ever before, the validity and reliability of an assessment program depends on the extent to which alignment considerations are integrated throughout the design process. Alignment considerations should begin with the choice of the Performance Expectation (PE) bundle and phenomenon and continue throughout each stage of development. In this paper we will describe guiding questions, strategies, and considerations for each phase of item development to aid the item developer in achieving well-aligned, multidimensional items.

INTRODUCTION

Assessment items developed against prior science standards or frameworks had focused heavily or solely on science domain content. In some cases, a portion of item pools were dedicated to aspects of inquiry or to the nature of science, but each was typically treated as a distinct construct relative to the core ideas and understandings within a particular scientific domain (Pellegrino, 2016). As a result, item alignments under prior frameworks were often unidimensional. In contrast, the NGSS and other K–12 Framework-inspired standards require a more integrative approach to construct definition—one that interweaves science practices, disciplinary core ideas, and crosscutting concepts. Therefore, item alignments supporting new assessments measuring three-dimensional science standards should be commensurately multidimensional, and each new assessment item must be carefully constructed to align with multiple dimensions of a particular PE or PEs. Given the complexity of multidimensional alignment claims, evaluation of such claims should be an integral component of each phase of assessment design and item development to ensure the validity of the final assessment. In addition, evidence to support alignment claims must be gathered throughout the process and carefully reviewed after completion and implementation of the assessment (Ruiz-Primo, DiBello, & Solano-Flores, 2014).

DESIGNING FOR ALIGNMENT

Beginning with the End in Mind

In NGSS-driven assessment design, strong item alignments are both an end goal and a design tool. The "backward design" process (Wiggins et al., 1998), which has been advocated for NGSS-aligned instructional models, curriculum frameworks, and formative assessments (e.g., Bybee, 2015; NRC, 2014; Whitehouse, 2014) with evidence based evaluations guiding the process, can also be applied to the development of NGSS-aligned large-scale assessment items. PEs and their associated foundation boxes provide learning outcomes, and thus serve as the logical starting point for a backward design model (Bybee, 2015). By beginning with the end in mind, alignment can be progressively strengthened throughout the development process.

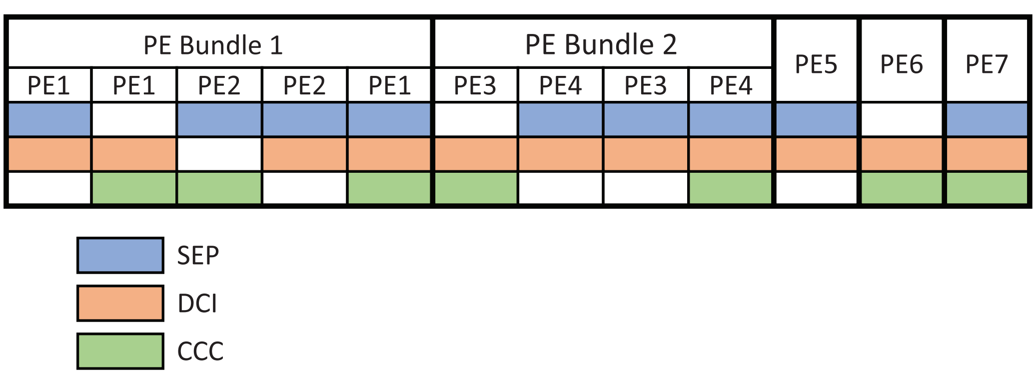
Item Alignments

Using the model in the *Assessment Framework* (CCSSO, 2015), an NGSS-aligned assessment typically comprises two distinct, but highly complementary, approaches to measurement:

1. Item clusters or other multi-item tasks, each of which aligns to a PE bundle (typically 2 or 3 PEs); and
2. Standalone (or discrete) items, each of which aligns to a single PE.

Whether considering item clusters or standalone items, the assessment item continues to serve as the basic unit of alignment. Item clusters achieve full coverage of all the dimensions of each PE within a PE bundle by mapping each item within the cluster against 2 or 3 dimensions of one or more of the PEs. For information on the process of developing an NGSS-aligned item cluster, please see the companion paper "Design and Development of Multidimensional Science Item Clusters," (CSAI, 2017a). For information on processes for maintaining quality while scaling up the development of item clusters, please see the companion paper "Quality Expectations and Development Considerations of Item Clusters Assessing Multidimensional Science Standards," (CSAI, 2017b). Standalone items assess 2 or 3 dimensions of a single PE, and are typically used to increase the breadth of PE coverage in an assessment. Figure 1 shows a sample alignment plan for an assessment that includes 7 PEs.

Figure 1: **Sample Assessment Alignment Plan**



Alignment Mapping

With respect to item cluster development, the PE bundle and the phenomenon used to examine the bundle provide the foundation of an alignment-driven design process. Alignment considerations are critically important even at this early stage of development, in part because the selection of a particular phenomenon and its association to a PE bundle represents the first alignment-driven decision in a backward design model. The SAIC Assessment Framework contains strategies for selecting a PE bundle that can enable efficient, focused assessment through the use of a single phenomenon presented in a stimulus. For standalone items, the backward design process begins with a detailed unpacking of a single PE (Harris et al., 2016).

Given the complexity and criticality of applying alignment considerations early on in the design process, assessment authors and practitioners may find it useful to leverage a specific set of guiding questions and principles. The following questions offer one such set of principles, in which each PE is examined through three complementary lenses. It is also worth noting that the three lenses described here can be applied in any order, but all three *must* be attended to in order to ensure strong alignment.

Table 1: **Questions to Guide Alignment-driven Design**

Lens	Overarching Question	Clarifying Questions
SEP	What does the PE language imply about how science and engineering practices (SEPs) should be used in conjunction with disciplinary core ideas (DCIs) and crosscutting concepts (CCCs) to explain the phenomenon?	▪ What does the language in the SEP dimension boxes, including the grade-band specific bullet, imply about evidence that could be used to determine student performance? ▪ Does the test design permit expansion of assessable limits to additional components of the practice listed in the grade-band SEP progression in Appendix F of the NGSS (NGSS Lead States, 2013) (? If so, what is the complete list of SEP components that could be drawn upon to gather evidence of student performance that is in alignment with the PE bundle? ▪ What types of stimuli will students need to interact with to demonstrate the application of the SEP(s) in the context of the phenomenon (e.g., data, evidence, models, claims, simulations)?

Lens	Overarching Question	Clarifying Questions
DCI	What does the language of the PE imply about how DCIs should be used in conjunction with SEPs and CCCs to explain the phenomenon?	■ What does the language in the DCI dimension boxes imply about the knowledge that students are expected to demonstrate as a result of instruction? ■ How do assessment boundaries, clarification statements, and DCI learning progressions in Appendix E of the NGSS (NGSS Lead States, 2013)limit the scope of knowledge that students are expected to demonstrate in the relevant grade band? ■ Given the core ideas students are expected to bring to the assessment, and limitations on the knowledge students can be expected to have, what information will need to be provided to students in the stimuli to fully explain the phenomenon? ■ How can questions about the phenomenon be constructed to not only elicit evidence of student knowledge of the DCI(s), but also to ensure that students are able to apply their knowledge in a novel context?

Lens	Overarching Question	Clarifying Questions
CCC	What does the PE language imply about how the CCCs should be applied in conjunction with DCIs and SEPs to explain the phenomenon?	▪ Does the test design require explicit evidence of students' ability to apply the crosscutting concept? If so, what does the language in the CCC dimension boxes imply about evidence that could be used to determine student CCC application? ▪ Does the test design permit expansion of assessable limits to additional bullets from the grade-band CCC progression in Appendix G of the NGSS (NGSS Lead States, 2013)? If so, what is the complete set of aspects of the CCC that could be drawn upon to gather evidence of student CCC application that is in alignment with the PE bundle? ▪ How can stimuli and items be crafted to support students' application of CCCs to explain the phenomenon?

Additional Considerations

If Evidence Statements are being used to support development, they can provide further insight into how the dimensions can be used in conjunction with one another in the examination of a phenomenon, as each Evidence Statement combines multiple dimensions. Evidence Statements can also help inform the types of evidence that can be elicited to evaluate student performance.

Choosing Phenomena and Stimuli – Data is Key

Alignment considerations applied during the first stage of analyzing PEs, selecting bundles, and choosing phenomena lead naturally to the next stage of development: crafting stimuli. Careful attention should be paid to the types of stimuli that allow students to use the SEPs to explain a phenomenon. Data, models, simulations, and other stimuli found in primary source materials may need to be modified and simplified for accessibility and grade-appropriate cognitive load. The importance of determining suitable data sources at this stage cannot be overemphasized (Addendum to the SAIC Assessment Framework, 2017). Without appropriate, accessible, relevant data, items are unlikely to achieve full alignment to most PEs. If sufficient data for a context cannot be found to craft stimuli aligned to the relevant PEs, a different context should be considered. More suitable contexts or phenomena may emerge from the process of searching for data. (A more detailed description of the process of item cluster development can be found in CSAI, 2017a.)

Coverage Mapping – A Planning and Evaluation Tool

In developing an item cluster, the next stage of the alignment process involves creating an outline to determine how items will target the range of dimensions necessary to achieve the required alignment to all PEs in the bundle. The outline should draw on draft stimuli to achieve a coherent arc for the student, including internal scaffolding as necessary. Drafting a cluster map is helpful at this point for checking dimensional coverage and for keeping track of any changes that may occur as items are developed. Figure 2 shows a sample map for an item cluster aligned to a two-PE bundle.

Figure 2: **Item Cluster Alignment Map**

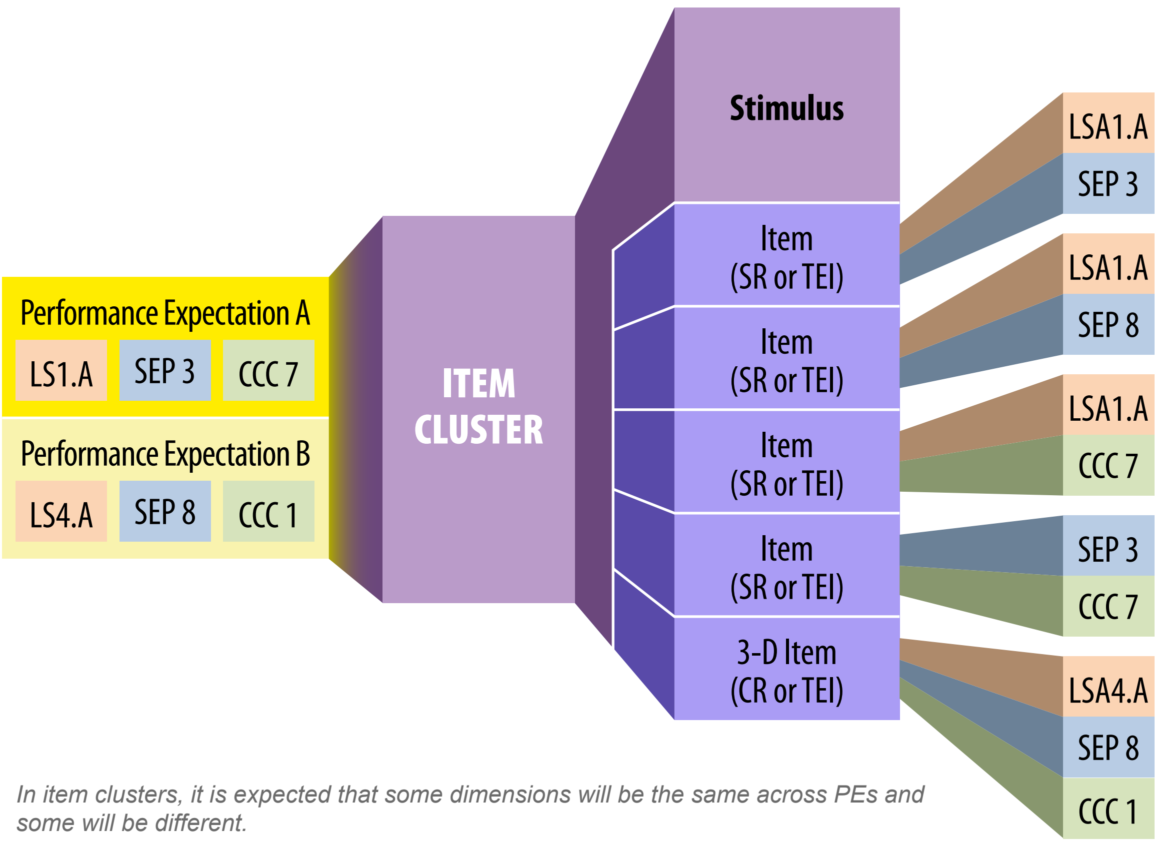


Image adapted from the [*SAIC Item Specifications Guidelines*]

CONSIDERATIONS DURING ITEM WRITING

Historically, item development has been considered an early stage of the design process. However, in a backward design model, writing items represents a late-stage effort—one that must be undertaken mindfully and deliberately, based on careful planning.

Multidimensionality

Whether within an item cluster or functioning as a standalone, the test design of most states requires that each item have a demonstrable and defensible explicit alignment to at least two of the three dimensions (SEP, DCI, or CCC) specific to the PE. Each item must support a strong enough alignment claim to protect the validity of a final assessment. Unidimensional items are too disparate from the integrative philosophy of the NGSS to be useful as a measurement tool, and must therefore be avoided.

Once a stimulus is drafted and the targeted dimensions for an alignment are chosen (i.e., mapped), items must be crafted in such a fashion that they can only be answered using knowledge and/or skills from the multiple dimensions they are intended to measure. For example, if a question can be answered only by analyzing a graph, without any specific knowledge of the DCI or application of the CCC, the item will not be considered two-dimensional, even if the data in the graph are from a context relevant to the DCI or CCC. The item must require the student to bring knowledge of the DCI or application of the CCC in order to complete the analysis. It is all too easy to oversimplify a context or edit work away from a two-dimensional alignment. When drifting occurs, developers may find it useful to refer back to the coverage map and the list of guiding questions that they employed at the outset of the design process. Examining the item through each dimensional lens often reveals alignment challenges.

Multipart Items

Items, even when functioning as discrete standalones, may merit subdivision into more than one part. However, each part of a given item should still strive to combine dimensions into a cohesive cognitive task as intended by the multidimensional approach in the SAIC Assessment Framework. Parts of an item should not be used to test ideas and/or skills in isolation. However, multipart items may be useful for scaffolding, supporting a claim or other response with evidence-based selected-responses (EBSRs), or elucidating an explicit connection to a crosscutting concept. Alignment judgments of multipart items should, however, be made on the composite of the item parts.

MEASURING MULTIDIMENSIONAL ALIGNMENT

Learning progressions, assessment boundaries, and Evidence Statements are all useful reference tools for the item developer to achieve the closest possible alignment to the targeted dimensions. Developers may also take guidance from tools that will be used for post-hoc evaluations, including alignment rubrics, models, and/or item specifications. Tools should contain criteria for judging dimensional alignments based on the item's ability to elicit student knowledge and/or skills required by multiple dimensions. Some tools elucidate these criteria in question form: if a student answers the item correctly, how confident is the reviewer in claiming that the student does indeed possess the knowledge and/or skills described in the relevant dimension boxes for the intended item alignment? Is it highly likely, likely, or impossible to tell? The lattermost evaluation indicates a lack of explicit alignment to the dimension in question.

Developing alignment evaluation tools in parallel with items creates challenges. In these situations, wherever possible, clarity should be sought around assessable limits (e.g., whether related elements of a learning progression are permissible) and implicit vs. explicit alignments (particularly relevant to the treatment of crosscutting concepts) prior to beginning development. At minimum, a systematic process of adjustment of the alignment evaluation tools and subsequent implementation should be clearly articulated. Table 1: Questions to Guide Alignment-driven Design beginning on page 5 can provide a useful starting point for clarification conversations.

CONTINUOUS EVALUATION AND CALIBRATION

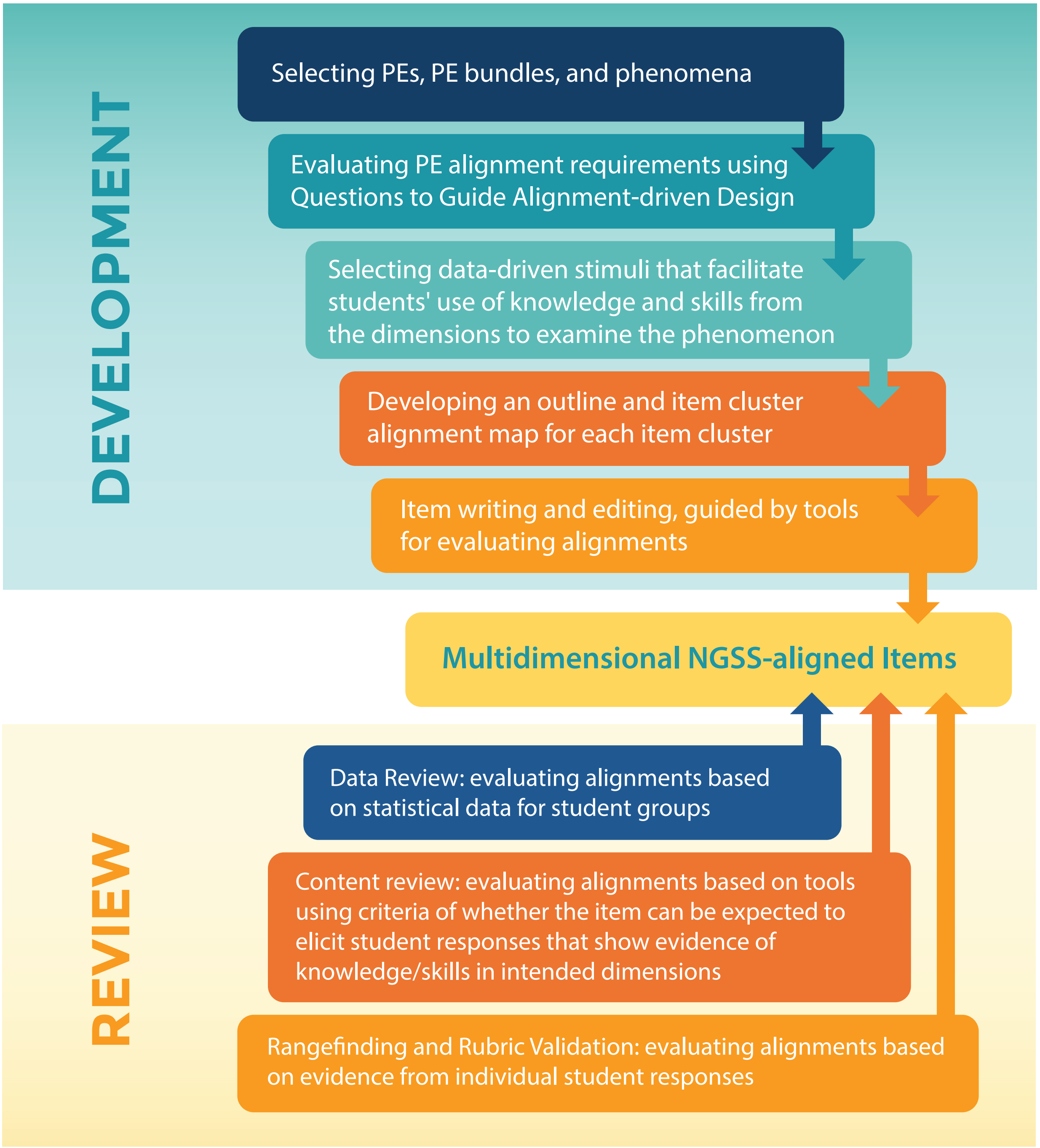
Each step in the development process provides an opportunity to review and evaluate the alignment of an item. Near the conclusion of initial development, a content review committee provides a critical "fresh eyes" step in fine-tuning the alignment of standalone items and item clusters to the dimensions of the PEs. With adequate training and the support of appropriate rubrics and tools for evaluating alignment, a content review committee can significantly contribute to the multidimensional alignment of the final version of the items. An important component of the tools used to evaluate alignment is the flexibility to indicate judgments of an item's degree of alignment to each dimension (e.g., full alignment to the SEP and partial alignment to the DCI). Each committee member's alignment judgments provide a starting place for rich discussions about the items and for targeted feedback and suggestions that lead to more fully aligned, multidimensional items.

The item development committees that come together after content review are typically prevented from making changes to items. Nevertheless, these committees have the advantage of item information that was not available to the earlier groups and can be called upon for an evaluation of item alignment. The rangefinding process allows alignment to be evaluated in light of evidence provided by student responses. These responses may reveal a weakness in alignment that was not apparent during content review. A final opportunity to judge item alignment prior to an item making its way into an operational item bank occurs at data review. Alignment issues discovered during rangefinding and data review should be used to inform choices made during the early stages of future development rounds.

CONCLUSION

The process of alignment-driven design described here incorporates a backward design process with evidence-based evaluations guiding many of the steps. Beginning with the end goal of alignment to multiple dimensions, as judged by the language of the PE, dimension boxes, and learning progressions, supports instruction and assessment in line with the holistic approach to multidimensionality elucidated in the K–12 Framework. Evaluating alignments using criteria based on whether student responses will provide evidence of alignments incorporates an evidence-based approach that further supports the holistic alignment philosophy. Outcomes of the evidence-based evaluations should be used as feedback for design steps as illustrated in Figure 3. In this manner, alignment for NGSS-based assessments can be approached as a continual improvement process.

Figure 3: **Designing for Alignment**



REFERENCES

Bybee, R. W. (2015). *The BSCS 5E Instructional Model: Creating Teachable Moments.* Arlington, Virginia: NSTA Press.

The Center on Standards and Assessment Implementation (CSAI). (2017a). *Design and Development of Multidimensional Science Item Clusters.* San Francisco, CA: The Center on Standards and Assessment Implementation

The Center on Standards and Assessment Implementation (CSAI). (2017b). *Quality Expectations and Development Considerations of Item Clusters Assessing Multidimensional Science Standards*. San Francisco, CA: The Center on Standards and Assessment Implementation.

Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & McElhaney, K. W. (2016). *Constructing Assessment Tasks that Blend Disciplinary Core Ideas, Crosscutting Concepts, and Science Practices for Classroom Formative Applications.* Menlo Park, CA: SRI International.

National Research Council (NRC). (2012). *A framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: National Academies Press.

National Research Council. (2014). *Developing Assessments for the Next Generation Science Standards.* Washington, DC: The National Academies.

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

Pellegrino, J. W. (2016). *21st Century Science Assessment: The Future Is Now.* (SRI Education White Paper). Menlo Park, CA: SRI International.

Penuel, W. R., Harris, C. J., & DeBarger, A. H. (2015). *Implementing the Next Generation Science Standards.* Phi Delta Kappan, 96(6), 45-49.

Ruiz-Primo, M. A., DiBello, L., & Solano-Flores, G. (2014). *Supporting the Implementation of the Next Generation Science Standards (NGSS) through Research: Assessment.* Retrieved from https://narst.org/ngsspapers/assessment.cfm

The Council of Chief State School Officers (CCSSO). (2015). *Science Assessment Item Collaborative (SAIC) Assessment Framework*. Washington, DC: Council of Chief State School Officers.

Whitehouse, M. (2014). *Using a Backward Design Approach to Embed Assessment in Teaching.* (SRI Education White Paper). Menlo Park, CA: SRI International.

Wiggins, G. P., McTighe, J., Kiernan, L. J., Frost, F., & Association for Supervision and Curriculum Development. (1998). *Understanding by Design.* Alexandria, Va: Association for Supervision and Curriculum Development.