

Structure of the NGSS

The Next Generation Science Standards (NGSS) are multidimensional with three dimensions:¹

- Science and engineering practices: 8 *practices* common to all disciplines of science and engineering.
- Cross-cutting concepts: 7 *concepts* common to all disciplines of science and engineering.
- Disciplinary core ideas: 45 *core ideas* in discipline-specific bodies of knowledge.

The *core ideas* are further divided into four disciplines:

- PS: Physical science
- LS: Life science
- ESS: Earth and Space Science
- ETS: Engineering, Technology, and Applications of Science

The NGSS uses performance expectations (or *expectations*) to describe science as carried out at the intersection of practices, concepts, and core ideas. Given the very large potential number of standards that could exist in the three-dimensional structure of the NGSS, using expectations also limits the content standards to a feasible number.

Clearing Up Potential Confusion

The Engineering, Technology, and Applications of Science (ETS) Discipline. The nature of the ETS discipline is ambiguous. Table 1 shows differences in how two parts of ETS are characterized in the National Academies of Science framework the NGSS is based on and in the NGSS². These differences cause some difficulty in summarizing the NGSS. I resolve this by treating ETS2.A and ETS2.B (See Table 1, below.) as core ideas, as this makes for clean figures and tables with minimal impact on results.

Table 1. *Different treatment of the ETS Discipline in the National Academies Framework and the NGSS.*

Structure of the ETS Discipline in the National Academies Framework	Appearance in the NGSS Performance Expectations
ETS1: Engineering Design	
ETS1.A: Defining and Delimiting an Engineering Problem	
ETS1.B: Developing Possible Solutions	
ETS1.C: Optimizing the Design Solution	Elements of core ideas
ETS2: Links Among Engineering, Technology, Science, and Society	
ETS2.A: Interdependence of Science, Engineering, and Technology	Closely tied with concepts
ETS2.B: Influence of Engineering, Technology, and Science on Society and the Natural World	(but not elements of concepts)

An Additional Component. The NGSS also includes the *nature of science* with eight elements. Four are closely associated with practices and four with concepts, as shown in Table 2. The nature of science is not another dimension, nor are its elements additional practices or concepts.³ I simplify by treating nature of science as composed of ancillary elements that co-occur with core ideas. To graphically represent NGSS’ structure, I developed figures and tables that incorporate all combination of 45 core ideas⁴, 8 practices, and 8 concepts⁵ and all combinations of 45 core ideas and 8 nature of science elements for a total of 3,240 intersections at which science

¹ See NGSS Lead States (2013a; 2013b).

² See Committee on a Conceptual Framework for New K-12 Science Education Standards (2012, p. 203) and NGSS Lead States (2013c)

³ See page 4 of NGSS Lead States (2013e).

⁴ Including ETS2.A and ETS2.B.

⁵ The seven listed *concepts* plus allowing an expectation to not list a *concept*. See expectation 3-5-ETS1-3 for an example (NGSS Lead States, 2013g).

may be practiced. I call this the “3D + 1” representation, which is demonstrated in Figure 1. Annotations in red curly braces (*{e.g.}*) show how elements of ETS2 and nature of science line up in the structure. The number of expectations covering a small subset of intersections is shown in Table 3 by grade.

Table 2. Treatment of the nature of science in the NGSS.

Elements of the Nature of Science	Appearance in the NGSS Performance Expectations
1. Scientific Investigations Use a Variety of Methods 2. Scientific Knowledge is Based on Empirical Evidence 3. Scientific Knowledge is Open to Revision in Light of New Evidence 4. Scientific Models, Laws, Mechanisms, and Theories Explain Natural Phenomena	Closely tied with practices (but not elements of practices)
5. Science is a Way of Knowing 6. Scientific Knowledge Assumes an Order and Consistency in Natural Systems 7. Science is a Human Endeavor 8. Science Addresses Questions About the Natural and Material World	Closely tied with concepts (but not elements of concepts)

Figure 1. Annotated excerpt of an expectation [see NGSS Lead States (2013f, p. 85)].

HS-ESS1-2		
Students who demonstrate understanding can: HS-ESS1-2. Construct an explanation of the Big Bang theory based on astronomical evidence of light spectra, motion of distant galaxies, and composition of matter in the universe. [Clarification Statement: Emphasis is on the astronomical evidence of the red shift of light from galaxies as an indication that the universe is currently expanding, the cosmic microwave background as the remnant radiation from the Big Bang, and the observed composition of ordinary matter of the universe, primarily found in stars and interstellar gases (from the spectra of electromagnetic radiation from stars), which matches that predicted by the Big Bang theory (3/4 hydrogen and 1/4 helium).]		
The performance expectation above was developed using the following elements from A Framework for K-12 Science Education		
Science and Engineering Practices	Disciplinary Core Ideas	Crosscutting Concepts
Constructing Explanations and Designing Solutions Constructing explanations and designing solutions in 9–12 builds on K–8 experiences and progresses to explanations and designs that are supported by multiple and independent student-generated sources of evidence consistent with scientific ideas, principles, and theories.	ESS1.A: The Universe and Its Stars <ul style="list-style-type: none"> The study of stars’ light spectra and brightness is used to identify compositional elements of stars, their movements, and their distances from Earth. The Big Bang theory is supported by observations of distant galaxies receding from our own, of the measured composition of stars and non-stellar gases, and of the maps of spectra of the primordial radiation (cosmic microwave background) that still fills the universe. Other than the hydrogen and helium formed at the time of the Big Bang, nuclear fusion within stars produces all atomic nuclei lighter than and including iron, and the process releases electromagnetic energy. Heavier elements are produced when certain massive stars achieve a supernova stage and explode. PS4.B: Electromagnetic Radiation <ul style="list-style-type: none"> Atoms of each element emit and absorb characteristic frequencies of light. These characteristics allow identification of the presence of an element, even in microscopic quantities. (secondary) 	Energy and Matter <ul style="list-style-type: none"> Energy cannot be created or destroyed—only moved between one place and another place, between objects and/or fields, or between systems.
Connections to Nature of Science {nature of science element 4} Science Models, Laws, Mechanisms, and Theories Explain Natural Phenomena {elements 1-4 appear with practices}		Connections to Engineering, Technology, and Applications of Science {core idea ETS2.A} Interdependence of Science, Engineering, and Technology {core ideas in ETS2 appear with concepts}
<ul style="list-style-type: none"> Construct an explanation based on valid and reliable evidence obtained from a variety of sources (including students’ own investigations, theories, simulations, peer review) and the assumption that theories and laws that describe the natural world operate today as they did in the past and will continue to do so in the future. 		<ul style="list-style-type: none"> Science and engineering complement each other in the cycle known as research and development (R&D). Many R&D projects may involve scientists, engineers, and others with wide ranges of expertise.
<ul style="list-style-type: none"> A scientific theory is a substantiated explanation of some aspect of the natural world, based on a body of facts that have been repeatedly confirmed through observation and experiment and the science community validates each theory before it is accepted. If new evidence is discovered that the theory does not accommodate, the theory is generally modified. 		<ul style="list-style-type: none"> Scientific knowledge is based on the assumption that natural laws operate today as they did in the past and they will continue to do so in the future. Science assumes the universe is a vast single system in which basic laws are consistent.

Finally, the expectations are combined into *topics*, covering each of the lowest-level groupings of core ideas. An example is provided in Figure 2, which shows the topic that contains HS-ESS1-2 (which was displayed in Figure 1).

Because the number of intersections of the dimensions of the NGSS covered by the expectations is small, some states are developing test content based on the complete set of intersections available in the topics, even though most of the intersections available in a topic are not covered by any individual expectation.⁶

Table 3. Number of expectations by grade or grade span.

Number of Expectations	Grade or Grade Span														
	Early Elementary			Late Elementary			Middle School			High School			Tested Grades	All Grades	
	K	1	2	3	4	5	6	7	8	9	10	11			12
Unique to a single grade	10	9	11	15	14	13	-	-	-	-	-	-	-	-	-
Shared by grade span	3			3			59			71			-	-	
Total by grade span	33			45			59			71			175	208	

Figure 2. Condensed excerpt from topic HS-ESS1 [see NGSS Lead States (2013c, pp. 97-98)].

HS-ESS1 Earth’s Place in the Universe		
Students who demonstrate understanding can:		
HS-ESS1-1. Develop a model based on evidence to illustrate the life span of the sun and the role of nuclear fusion in the sun’s core to release energy that eventually reaches Earth in the form of radiation.		
HS-ESS1-2. Construct an explanation of the Big Bang theory based on astronomical evidence of light spectra, motion of distant galaxies, and composition of matter in the universe.		
HS-ESS1-3. Communicate scientific ideas about the way stars, over their life cycle, produce elements.		
HS-ESS1-4. Use mathematical or computational representations to predict the motion of orbiting objects in the solar system.		
HS-ESS1-5. Evaluate evidence of past and current movements of continental and oceanic crust and the theory of plate tectonics to explain the ages of crustal rocks.		
HS-ESS1-6. Apply scientific reasoning and evidence from ancient Earth materials, meteorites, and other planetary surfaces to construct an account of Earth’s formation and early history.		
The performance expectation above was developed using the following elements from A Framework for K-12 Science Education		
Science and Engineering Practices	Disciplinary Core Ideas	Crosscutting Concepts
<ul style="list-style-type: none"> Developing and Using Models Using Mathematical and Computational Thinking Constructing Explanations and Designing Solutions Engaging in Argument from Evidence Obtaining, Evaluating, and Communicating Information 	ESS1.A: The Universe and Its Stars ESS1.B: Earth and the Solar System ESS1.C: The History of Planet Earth ESS2.B: Plate Tectonics and Large-Scale System Interactions PS1.C: Nuclear Processes PS3.D: Energy in Chemical Processes and Everyday Life PS4.B: Electromagnetic Radiation	<ul style="list-style-type: none"> Patterns Scale, Proportion, and Quantity Energy and Matter Stability and Change <p>Connections to Engineering, Technology, and Applications of Science</p> <ul style="list-style-type: none"> Interdependence of Science, Engineering, and Technology <p>Connections to Nature of Science</p> <ul style="list-style-type: none"> Scientific Knowledge Assumes an Order and Consistency in Natural Systems
<p>Connections to Nature of Science</p> <ul style="list-style-type: none"> Science Models, Laws, Mechanisms, and Theories Explain Natural Phenomena. 		

Domain Coverage in the NGSS

A few states administer science assessments in contiguous grades⁷, but most administer by grade span, so I use grade spans for tables and figures. Copying NGSS’ color scheme for figures, orange is for disciplinary core ideas (DCI), blue for science and engineering practices (SEP), and green for cross-cutting concepts (CCC). I add purple for nature of science (NOS).

The complete domain of science defined by the NGSS is the 3,240 cells per grade/span, but the expectations cover only a small subset. Table 4 summarizes this coverage by grade span. Figures 3 and 4 show degrees of coverage (by expectations and topics, respectively) in the 3-5 grade span. Figures 5 and 6 do the same for the middle school grade

⁶ As noted by state staff from Michigan and Delaware at the 2017 (September) Reidy Interactive Lecture Series in Portsmouth, NH.

⁷ See Alabama State Department of Education (2016), Louisiana Department of Education (n.d.), and Utah State Board of Education (2017).

span, and Figures 7 and 8 for the high school grade span. In each figure, the number of expectations or topics that address each cell in the domain matrix is shown in the cell. While text in the figures is small, the figures remain useful in understanding the challenges by depicting the very sparse coverage of the complete domain. An analogous table including each grade and grade span and corresponding, larger, more readable figures are given in Appendix A.

Table 4 shows that the total domain is poorly represented by expectations in all grade spans (the last two rows of the table). The coverage is less than three percent in grades 3-5, less than four percent in middle school, and less than five percent in high school. Likewise, the total domain is only somewhat better represented by topics in all grade spans (the middle two rows of the table). The topic-based coverage is less than seven percent in grades 3-5, less than twenty-five percent in middle school, and less than 30 percent in high school. Considered in isolation, coverage of core ideas, practices, concepts, and the nature of science is considerably better (the first four rows). However, even here there are some potential concerns: coverage of core ideas and the nature of science is incomplete in all three grade spans, and coverage of concepts is incomplete in the 3-5 grade span.

Sparseness of coverage is exacerbated with grade-level because in grade-span testing the coverage can accumulate over multiple grade levels. The degree of exacerbation with grade-level testing can be seen in Appendix A.

Table 4. Coverage of the NGSS by tested grade span⁸.

Dimension or Component	Type	Code	Percent of Elements of a Dimension or Component Addressed in Expectations (or in Topics)		
			Elementary (grades 3-5)	Middle School (grades 6-8)	High School (grades 9-12)
Disciplinary Core Ideas	Dimension	DCI	91	89	93
Science & Engineering Practices	Dimension	SEP	100	100	100
Crosscutting Concepts	Dimension	CCC	71	100	100
Nature of Science	Component	NOS	75	75	88
Combination of Dimensions and/or Components		Part of Domain	Percent of Domain Cells Addressed in Topics		
			3-5	MS	HS
DCI × SEP × CCC		3D	6.6	24.4	28.3
(DCI × SEP × CCC) + (DCI × NOS)		3D + 1	6.8	24.1	27.9
Combination of Dimensions and/or Components		Part of Domain	Percent of Domain Cells Addressed in Expectations		
			3-5	MS	HS
DCI × SEP × CCC		3D	2.3	3.2	3.8
(DCI × SEP × CCC) + (DCI × NOS)		3D + 1	2.5	3.7	4.5

Implied Nested Data Structures

The “3D + 1” structure of the NGSS sends educators a specific signal. That signal is that practices, core ideas, concepts, and the nature of science should be integrated in instruction because science is practiced at their intersection. That is, it is not possible to successfully practice, for example, life science without needing an understanding of the practices and concepts. Therefore, faithful assessment of the NGSS is unlikely with discrete items, and more likely with *item sets* built around a bundle of potentially interactive stimulus material to represent the “3D + 1” structure. This creates nested data structures, for example, with items nested within item sets.

⁸ Thanks to Jon Cohen of American Institutes for Research who noted on an earlier version that I was working from *topics* [see NGSS Lead States (2013c)] rather than *expectations* [see NGSS Lead States (2013f; 2013g; 2013h)]. The issues identified are considerably greater when working from *expectations*. *Topics* include many more combinations of dimensions than do *expectations*. However, it is *expectations* (not *topics*) that represent content eligible for inclusion on assessment.

A Conflict Arising from the Structure and Coverage of the NGSS

Two conflicting considerations that arise with this set of large and sparse matrices are (1) *focus*, and (2) *transfer* (with a corollary concern of narrowing the curriculum)⁹. The issue of focus arises because without some method of limiting the scope of curriculum, instruction, and assessment, the complete domain can be overwhelming. As was done with the NGSS, developing expectations covering only select cells of the domain helps to mitigate this issue.

The consideration of transfer arises because limiting the scope of curriculum, instruction, and assessment to a feasible subset of the full domain can lead to concerns about whether students can translate the practices, crosscutting concepts, and understanding of the nature of science to phenomena not covered in the core ideas included in curriculum and instruction. Evaluating this concern would require introducing a few new expectations covering additional cells with associated item sets and comparing student performance on new and existing sets.

Implications for Selecting a Measurement Model

Nested Data Structures

The use of item sets strongly tied to a specific scenario and/or a set of stimulus material is likely to produce an appreciable degree of dependency among item scores, causing some problems for widely used item response theory (IRT) models.¹⁰ In addition, the nesting is more complex than it may appear at first blush. Items will likely be nested in an item set associated with an expectation that is based on a scenario and/or an interactive stimulus. Nested within an item set, there may be *item trees*, defined as either

- Items administered only in the case of specific responses to one or more previous items, or
- Items that can only be scored in light of responses to one or more previous items.

Finally, because the expectations envision knowledge integrated across dimensions, methods for scoring student responses may also introduce additional nesting. For example, examinee interactions with the stimulus material and/or multiple items may be considered in creating scores associated with one or more interactions with one or more items and/or stimuli. American Institutes for Research (AIR) calls these *scoring assertions*.¹¹ With scoring assertions, the number of scores produced for an item set may be greater or lesser than the number of items in the set, and any specific score may be cross-nested within an item set, item tree, and even multiple individual items. Item sets, item trees, and scoring assertions can create dependencies among scores that may need to be accounted for in the measurement model. However, other approaches are also possible.¹² It may be that a simple, traditional model is robust to violations of local item dependence (meaning that the violations have minimal impact on results). In that case, the complexity of these extensions of item response theory (IRT) can be avoided. Another approach is to sum each set of dependent scores and treat each summed score as a single item (but this approach may hide problems rather than address them).

Dimensionality

A dimension can be conceptually described as an underlying coherent group of knowledge and skill within a content area in which it would be difficult to defend subdividing the group into smaller pieces using both substantive and statistical criteria.¹³ In educational assessment, dimensions are typically conceptualized as sub-domains of a content area, such as discipline (e.g., earth and space science, life science, and the physical science groupings of the NGSS core ideas). They are not typically conceptualized as aspects that cut across disciplines within a content area (e.g.,

⁹ This concern was prompted for me by a presentation by Jim Pellegrino (2017).

¹⁰ See Andrich (1983); Zenisky, Hambleton, and Sireci (2002); and Sideridis (2011).

¹¹ See Doran (2017) and Rijmen (2017).

¹² See Cao, Lu, and Tao (2014).

¹³ See Li, Jiao, and Lissitz (2012).

practices, concepts, or nature of science groupings in the NGSS). This may create some confusion about dimensionality because the term “dimension” is used differently in the NGSS.¹⁴

However, that does not mean that the dimensions as defined in the NGSS will not function as distinct psychometric dimensions (even though integration across all aspects of the NGSS is intended in curriculum, instruction, and assessment). Since we do not yet have much experience with well-designed NGSS assessments, it is not clear whether these assessments will tend toward essential unidimensionality in the way other large-scale content area assessments have.¹⁵ There are various possibilities for dimensional structure in NGSS-based assessments, including but not limited to the following:

- A complex structure with seven dimensions: physical science (PS) core ideas, life science (LS) core ideas; earth and space science (ESS) core ideas; engineering, technology, applications of science (ETS) core ideas; practices, concepts, and nature of science (with the last three cutting across the first four).
- A less complex structure with five dimensions: PS core ideas, LS core ideas, ESS core ideas, ETS core ideas, and general science (i.e., a conglomeration of practices, concepts, and nature of science).
- A still less complex structure with four dimensions: PS core ideas, LS core ideas, ESS core ideas, and general science (i.e., a conglomeration of ETS core ideas, practices, concepts, and nature of science).
- A unidimensional structure in which no dimensions are separable.

Unmodeled dimensionality can have a negative effect on score validity similar to the effects of unmodeled nested data structures: by creating local item dependencies.¹⁶ It is important to evaluate whether scores are “essentially unidimensional enough” to be robust to local item dependence.¹⁷ If scores are not essentially unidimensional enough, it will be important for content area experts and psychometricians to identify the appropriate dimensions to include in the measurement model. It is important to note that the number of dimensions to report on may justifiably differ from the number of dimensions in the measurement model.

Selecting a Measurement Model

Addressing Nested Data Structures. While there is considerable literature on IRT extensions that account for nesting, I am not aware of any state testing program that has implemented them in operation. If it is necessary to use extensions of IRT that account for nested data structures, it will likely be a new practice for most vendors working with states. It will also present challenges for states in exercising psychometric oversight of vendors and in transitioning from one vendor to another if the new vendor is unfamiliar with the model.

Given the complexity of this issue, states should leverage psychometric expertise of their technical advisors (and in-house staff, if available) to work with vendors’ psychometric staff to address this issue. States will need to evaluate the need for more complex measurement models. It is possible that widely used, simple measurement models are robust to the potential violations of local dependence introduced by nested data structures. Existing research suggests using model fit indices to evaluate the appropriateness of the following¹⁸:

- Simple measurement models without modification;

¹⁴ Thank you to Nathan Dadey and Scott Marion of the Center for Assessment for pointing this out in response to a first draft of this paper.

¹⁵ See Nandakumar (1991).

¹⁶ This occurs because item scores are assumed to be uncorrelated after accounting for estimates of latent achievement. If too few dimensions of achievement are included in the model, items affected by the same unmodeled dimension of achievement will remain correlated even after accounting for estimated achievement.

¹⁷ See Reise et al. (2013) and Li, Jiao, and Lissitz (2012) for examples.

¹⁸ Jiao and Zhang (2017) indicate that in a complex cross-nested data structure, the deviance information criterion (DIC) performed better than the Akaike (AIC) and Bayesian (BIC) information criterion (a logical conclusion given that DIC is a hierarchical generalization of AIC and BIC).

- Simple measurement models on summed dependent scores treated as a single item (not generally recommended¹⁹); or
- IRT models that explicitly address data nesting²⁰, such as
 - Explicitly-testlet-based models²¹,
 - Mixture models²²,
 - Hierarchical models²³,
 - Partitioning models with a primary construct-relevant dimension and a nuisance dimension²⁴, or
 - Some other novel approach.

Of particular interest are some timely, preliminary results shared by AIR regarding the cluster-based NGSS assessment it is developing with various states. These results indicate, as anticipated, that at least in the AIR assessment, nesting structures introduce considerable noise that must be accounted for.²⁵

To address these issues, technical deliberations should address the fit of the measurement model to the structure of standards and design of the assessment, as evaluated and negotiated among content and psychometric experts.²⁶ The theoretical basis for selecting a measurement model should be balanced with empirical analyses of the robustness of less complex models. States' technical advisory committees (TACs) can be helpful in reviewing evidence from vendors' and/or states' psychometric staff.

Since this will be new to both states and vendors, it will be important for the TAC to provide feedback on plans to ensure they are thorough. If states do not have TAC members with expertise in complex models that account for nested data structures, it will be important to identify someone with that expertise.

This will be needed to ensure that vendor plans and completed work are conducted appropriately and that the theoretical and empirical basis for model selection is soundly documented. It will be important to thoroughly document the methods of implementation to ensure the program can survive staff turnover or transition to a new vendor. The documentation must be sufficiently detailed to allow for replication of analyses. The software, the code used to run the software, data cleaning rules, and business rules for data manipulation must also be thoroughly documented to make replication possible. A state's TAC should review the documentation to ensure it can be replicated

Addressing Multidimensionality. It is doubtful whether it is desirable to report dimension subscores ((to discourage isolating parts of the integrated “3D + 1” structure in instruction). At the same time, because the standards are multidimensional in a way not yet seen in other academic content standards, it may be important to consider dimensionality in selection of a measurement model even if dimensionality is not reported on.²⁷ As with nested data structures, there is also considerable literature on evaluating whether data are essentially unidimensional enough to justify a unidimensional measurement model.²⁸ States should conduct analyses when initially calibrating the assessment. If data reasonably support a unidimensional model upon initial calibration, the analyses should be repeated periodically to evaluate whether that model remains appropriate. If there is a need to report on multiple

¹⁹ While this is a commonly used option, it can “hide” multidimensionality in the summed item scores (Reise, Scheines, Widaman, & Haviland, 2013). This is likely to be inappropriate for NGSS assessments that are faithful to the intended “3D + 1” structure.

²⁰ Note that the overlap between these categories is considerable.

²¹ See Li, Bolt, and Fu (2006).

²² See Jiao and Lissitz (2015) and Mislevy, et al. (2007).

²³ See Bolt and Kim (2015) and Jiao and Zhang (2017).

²⁴ See Doran (2017) and Rijmen (2017).

²⁵ See Cohen (Cohen, 2017).

²⁶ See Mislevy, et al. (2007) for a principled theoretical approach to selecting a measurement model.

²⁷ See Dorans and Kingston (1985); Camilli, Wang, and Fesq (1995); De Champlain (1996); Cao (2008), and Bolt (1999) for a body of research showing contrasting results of both minimal and consequential effects of unmodeled dimensionality on equating.

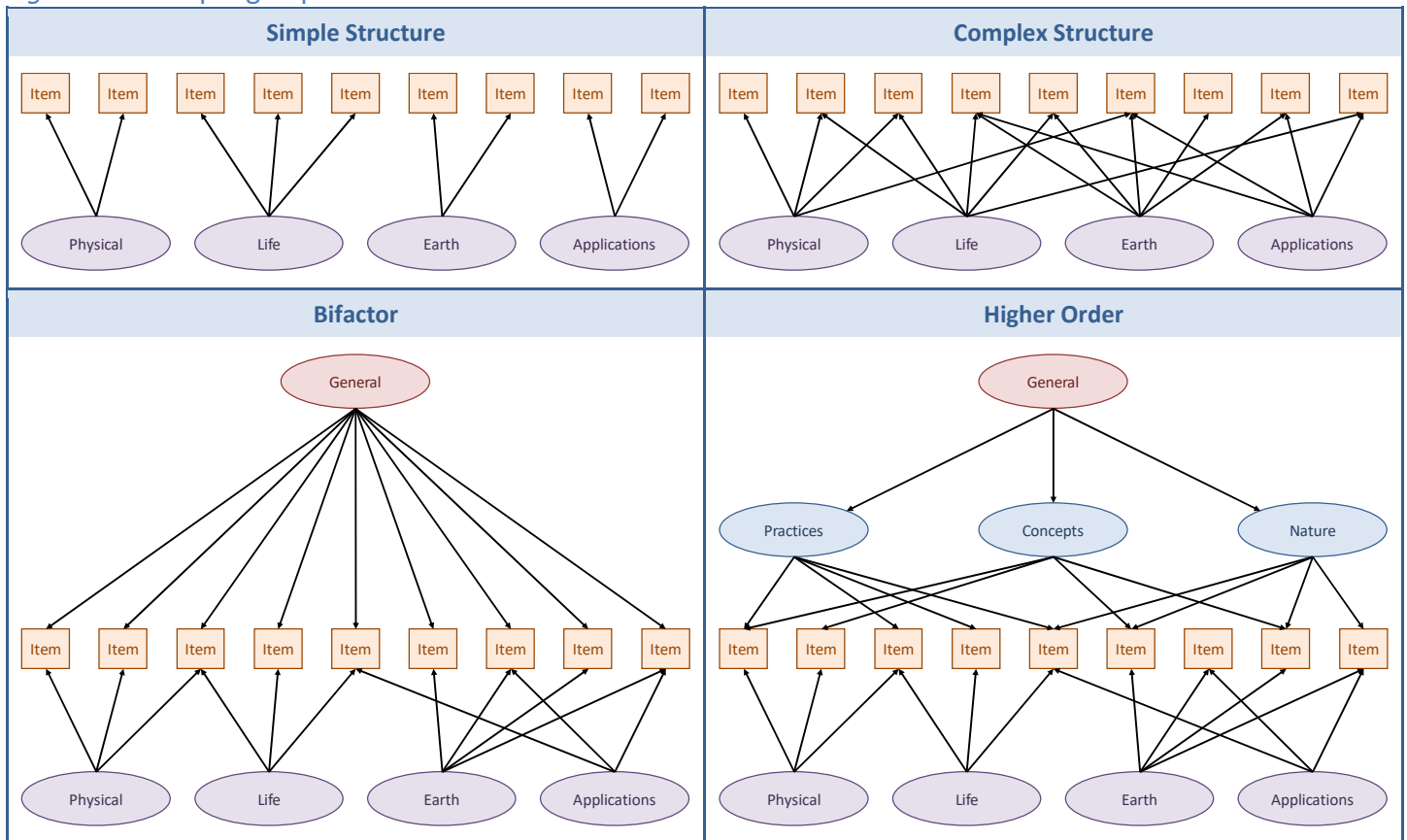
²⁸ See Reise, et al. (2013) for a relatively recent high-level summary and Zeng (2010) for a more dated but more thorough review.

dimensions even though data support a unidimensional model, it may be reasonable to use item parameters from an overall calibration to estimate dimension scores. But this should be checked initially and over time.²⁹

On the other hand, if initial calibration data do not reasonably support a unidimensional model, the process of selecting a measurement model becomes more difficult. Since there apparently is not a considerable body of successful work defining factors using exploratory factor analysis (EFA) in K-12 educational assessment, I recommend another approach. Vendor and state technical staff may be better served if they work with subject matter experts. They can help to define one or more theoretical factor structure based on substantive knowledge. The resulting structure(s) can then be analyzed using confirmatory factor analysis (CFA). While EFA has generally been unsuccessful in *defining* dimensions, EFA approaches can be useful in one regard when scores are not essentially unidimensional. They can be helpful in *identifying the approximate number* of dimensions, but only if certain criteria are used. This is because many commonly-used criteria perform poorly when dimensions are correlated (as is generally true with educational assessment).³⁰ There are many potential models that can be used for multidimensional score data. In the CFA analyses performed to select a measurement model, knowing the approximate number of dimensions in the data can be helpful.

A variety of factor structures can be represented in a measurement model. In Figure 9, item scores are shown in tan, dimensions based on disciplinary core ideas in purple, aspects that cross disciplines (practices, concepts, and nature of science) in blue, and general science knowledge in red.

Figure 9. A sampling of potential factor structures and associated measurement models



²⁹ See Martineau and Zeng (2015) for an example of such an evaluation.

³⁰ Of measures in existence at the time, Zeng (2010) found three uncommonly used measures (maximum likelihood chi-square; Akaike information criterion, or AIC; and Bayesian information criteria, or BIC) to perform well whereas many traditionally used measures performed poorly (e.g., scree test, eigenvalue ratios, parallel analysis). Likewise, in the context of nested item structures as unmodeled dimensionality, Jiao and Zhang (2017) found that DIC (deviance information criterion) performed better than either AIC or BIC (a logical result as DIC is a hierarchical generalization of AIC and BIC).

The or arrows show which elements of the model affect others. For example, items (in rectangles) always have arrows pointing inward because knowledge (in ovals) affects how students respond rather than the other way around. In the *simple structure* panel, each item score is affected by knowledge about only a single discipline. In the *complex structure* panel, each item score is affected by knowledge in at least one discipline, but may be affected by multiple disciplines. In the *bifactor* panel, there is a general science dimension representing all aspects of the NGSS that appear across disciplines, with specific disciplinary factors. Responses to specific items are affected by both the general science dimension and one or more specific disciplinary dimensions. Finally, in the *higher order* panel, there is a general science dimension, but there are also specific dimensions for practices, concepts, and nature of science.

As with nested data structures, I am not aware of any state testing programs that use a multidimensional measurement model currently in operation, and only one consortium of states.³¹ If it is necessary to use extensions of IRT that account for nested data structures, it will likely be a new practice for most vendors. It will also present challenges for states in exercising psychometric oversight of vendors and in making the transition from one vendor to another if the subsequent vendor is unfamiliar with the model. Again, states should leverage psychometric expertise of their technical advisors (and in-house staff, if available) to work with vendors' psychometricians to address this issue. States will need to evaluate the need for more complex measurement models. Again, it is possible that widely used, unidimensional measurement models are robust to potential violations because the data are "essentially unidimensional enough." Existing research again recommends the use of model fit indices in determining essential unidimensionality, and, if necessary, an appropriate multidimensional model.

Analyses used to determine whether the measurement model should be unidimensional or what type of multidimensional model should be used should include the following steps:

- Evaluation of whether the assessment is "essentially unidimensional enough" to support the use of a unidimensional model.³²
- If the data support the use of a unidimensional model:
 - Work with content area experts to identify any needed reporting of dimension scores. In general, it is better to have fewer levels of scores (e.g., one level of subscores, such as *domains* vs. two levels of subscores, such as *domains* and *targets*.) This is because subscores will likely provide little additional information (only identifying exceptional cases as having an informative score profile), and sub-scores are unlikely to provide any useful information.³³
 - Evaluate whether the dimension scores can be calculated from unidimensional item parameters³⁴ or whether they will need to be recalibrated using data from only the items identified as part of a specific dimension.
 - If the data support using the unidimensional item parameters, plan periodic re-evaluations of whether the data still support this use.
- If the data do not support the use of a unidimensional model:
 - Content area experts and psychometric staff should work together to identify the theoretically ideal dimensionality structure based on available models.
 - Using confirmatory factor analysis (CFA) and model fit indices, the psychometric staff and content area experts should trim the ideal model to one that is both theoretically useful and empirically supported.³⁵
 - The content area experts and psychometric staff should jointly determine what dimension should be reported, if any. Modeling scores as multidimensional does not necessitate reporting scores on all dimensions modeled. There may be strong substantive reasons for reporting scores on either fewer or more dimensions than are modeled.

³¹ The ELPA21 Consortium uses a multidimensional IRT model (personal communication with Jon Cohen, September 2017).

³² See Reise, et al. (2013) for a relatively recent high-level summary and Zeng (2010) for a more dated but more thorough review.

³³ See Monaghan (2006) for a conceptual explanation of issues with subscores and criteria for evaluating their usefulness, with references to more technical papers.

³⁴ See Martineau and Zeng (2015) for an example of such an evaluation.

³⁵ See Li, Jiao, and Lissitz (2012), Reise (2012), and Reise et al. (2013) for examples.

- The structure of the model should be evaluated on a periodic basis as the program matures.

Again, given the complexity of initial analyses and ongoing monitoring, I recommend that vendor (and state) staff bring proposed analyses and models with supporting evidence to the state's TAC to develop and refine a plan for determining the dimensionality of the data and to review and provide feedback on results in an iterative fashion.

In both cases (nested data and multidimensionality) the newness of more complex models means that there are two additional proactive steps that may be helpful. First, if the field test is carried out in a manner that strongly resembles what the operational test will be like, model selection can reasonably be performed using field test data. Otherwise, it is better to wait until operational data are available, even if it means an additional delay in reporting. This is because if a model is selected using field test data but the model does not fit operational data, the delays will be considerably longer.

Second, it will be important for states to have confidence that the implementation of a more complex model is adequately documented. A novel psychometric model may not have standard software or rules of thumb yet developed. Turnover in vendor staff or a transition from one vendor to the next could introduce serious problems if the model is implemented differently across years. Therefore, highly detailed business rules will be necessary, and should be evaluated by an external expert with knowledge of the models being used. Consistency in implementation will be key to smooth transitions across staff and vendors. TACs can be helpful in evaluating the completeness of documentation.

Combined Approach to Addressing Nested Data Structures and Dimensionality

Nested data structures often manifest as dimensions, so it will likely be possible to combine the approaches to addressing the two issues. For example, it may be that a model like the one shown in Figure 10 is appropriate for NGSS assessment data. This figure is shown as a hypothetical example that could be useful.

Panel 1 depicts a typical unidimensional IRT model. Panel 2 depicts discipline-specific dimensions (e.g., general science knowledge may be insufficient to respond to some tasks in, say, life science). Panel 3 depicts added nuisance dimensions from nesting of scoring assertions within an item set. Finally, panel 4 depicts the addition of additional nuisance dimensions for multiple scoring assertions nested within an individual task or item.

While it is unlikely that each type of dimensions will be necessary, it will be important to evaluate which will be necessary.

Implications for Equating

Nested Data Structures

The use of models that account for local item dependence introduces additional parameters into the equating structure. Equating can still be done, but may require specialized methods. Options include using a specialized equating method developed for the model (few novel models will have a dedicated equating method) or number correct or true score equating.³⁶ However, some more complex models may have important desirable theoretical characteristics that are lost when scores are equated using number correct or true score equating.

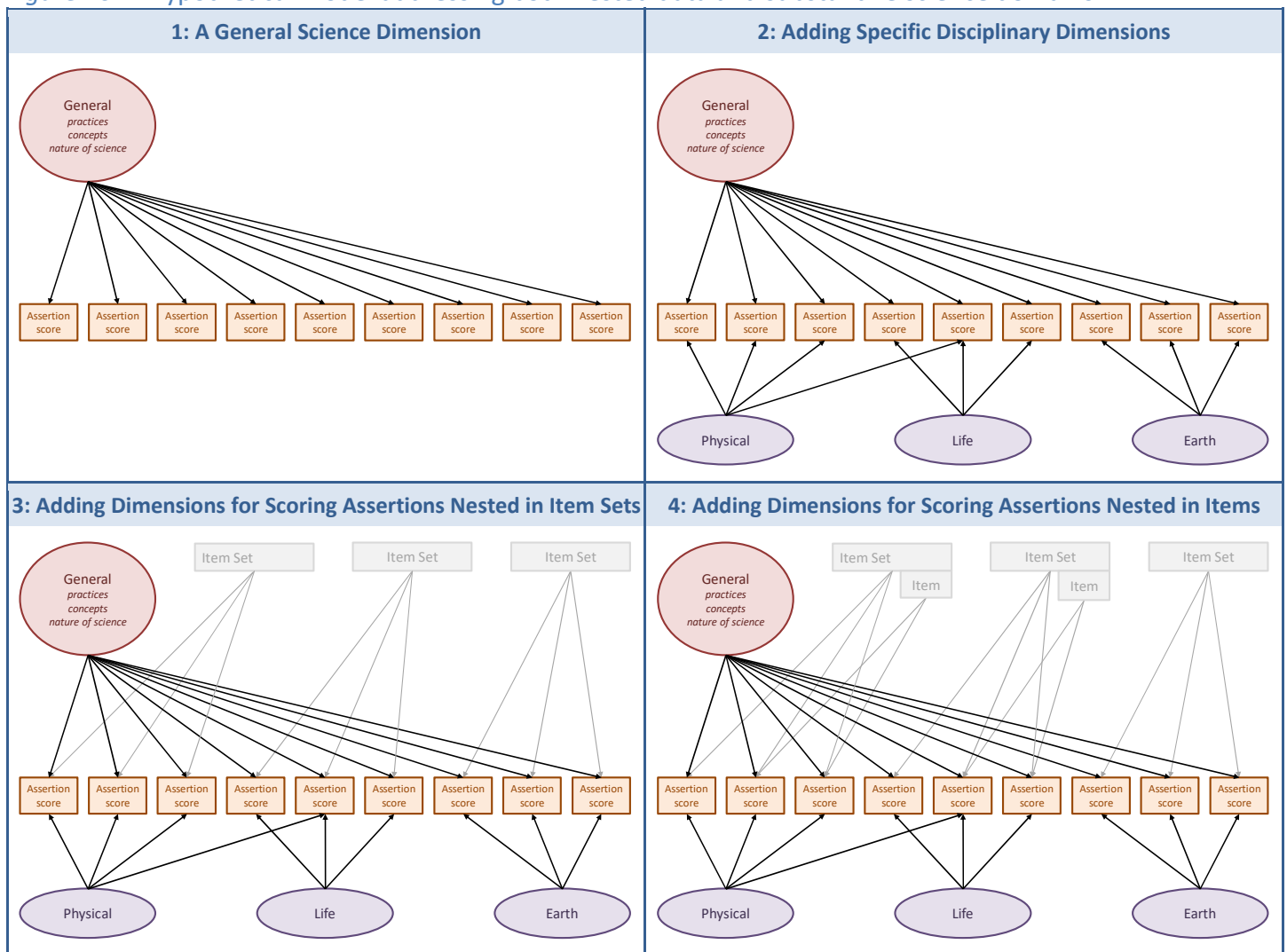
Sparse Domain Coverage

The implications for equating depend on how the conflicting issues of focus and transfer are balanced. If focus is prioritized at the expense of transfer, assessments will be limited to the current set of expectations and equating-

³⁶ See Chen (2014) and Cao, Lu, and Tao (2014) for examples.

specific implications would be minimal as the limited sample of the domain covered by the assessment will not change considerably over time. On the other hand, if transfer is prioritized at the expense of focus, the entire domain matrix would be eligible for sampling and new expectations would be developed to cover matrix cells not currently included. The implications for equating could be considerable in that as new expectations are developed a broader domain is available for sampling. If transfer is poor or is differentially manifested for different sub-populations of students, considerable differential item or test functioning could occur, causing sample-dependency in equating functions. Monitoring for such sample dependencies would be a critical task. This can be done by identifying sub-populations with enough students that can be monitored during equating.

Figure 10. A hypothetical model addressing both nested data and substantive science domains.



It may be possible to balance the priorities of focus and transfer. One way to do this would be to treat the entire matrix as the domain but treat the existing expectations as a core set of standards that is measured on every test form. The remaining cells of the matrix could be sampled each year to develop new expectations and associated test content for later test forms. In this way, existing expectations represent a non-random but strategically-selected core with new expectations representing random sampling from the remainder of the domain to check for transfer. By using this strategy, equating items would always be drawn from the existing expectations to improve equating stability over time.

Finally, sparse domain coverage introduces concerns about stability in equating if a substantial set of standards is not included on test blueprints every year. This may necessitate creativity in expanding the linking item set as described below.

Dimensionality

While there is considerable literature on multidimensional equating, generalizing unidimensional equating methods to the multidimensional case³⁷, I am not aware of state tests that have successfully implemented multidimensional equating in practice. This will likely be a new practice for most vendors working with states. Again, states should leverage the psychometric expertise of their technical advisors and/or in-house staff to develop a plan to evaluate the appropriate methods for equating based on the measurement model selected. As with nested data structures, options include using a specialized equating method developed for the model (some multidimensional models, especially novel ones, will not have specialized methods) or using number correct or true score equating.³⁸ However, some more complex measurement models may have important, desirable theoretical characteristics that are lost when scores are equated using those methods.

An additional implication of multidimensionality is the need for larger linking item sets. Along with sparse domain coverage, this may require some creativity in expanding the linking item set.

Selecting an Approach to Equating

Addressing Nested Data Structures and Multidimensionality. The implications for equating depend on selection of a measurement model. I recommend that when vendors and states consider issues of nested data structures and multidimensionality in measurement model selection, they also consider downstream effects on equating. These issues should also be taken to the state's TAC alongside considerations for measurement model selection.

Addressing Sparse Domain Coverage. The key step in addressing sparse domain coverage is to determine the relative weight of the conflicting issues of focus and transfer. This is primarily a policy-based decision, and should be led by the state's content and policy staff. However, technical staff should be included in discussions so that content and policy staff are made aware of any decisions with technical ramifications being considered.

If the decision is to privilege focus (i.e., treating the existing expectations as the full set of cells from the domain eligible to appear on the assessment), this issue is moot. On the other hand, any other decision will require careful consideration and involvement of technical advisors.

If the decision is to privilege transfer (i.e., treating the full matrix of the "3D + 1" domain of the NGSS as eligible to appear on each form of the assessment), it will be important to consider how the equating design will ensure that there is sufficient similarity in horizontal and vertical linking item sets. This is important across all forms both within and across years, including the likelihood that sparse domain coverage may increase the likelihood of scale drift. These considerations should be brought to the state's TAC to review and to provide feedback in an iterative fashion.

Finally, if the decision is to balance focus and transfer (i.e., treating the existing expectations as a core that is eligible to appear on every form of the assessment, with novel expectations written to cover other cells of the domain matrix eligible for inclusion on a matrix sampling basis), the issues are less concerning. The core of expectations eligible to appear on every form of the assessment can help to maintain consistency of linking item sets across forms. Equating methods, analyses of horizontal and vertical scale drift, and analyses comparing performance on

³⁷ For example, see Hirsch (1989); Brossman and Lee (2013); Davey, Oshima, and Lee (1996); and Chen (Chen, 2014).

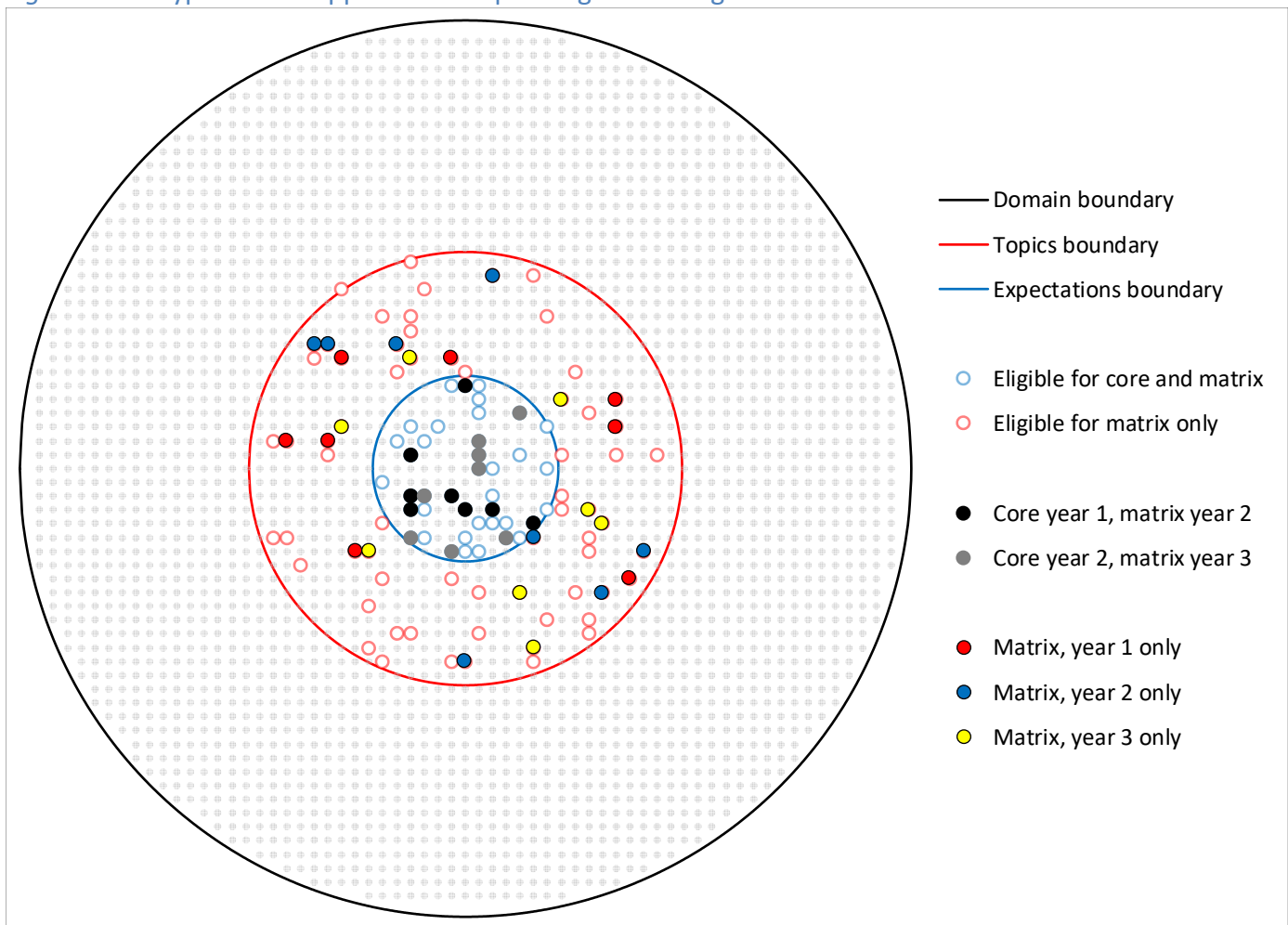
³⁸ See Chen (2014) and Cao, Lu, and Tao (2014) for examples.

the core (existing) expectations and matrix sampled (new) expectations will need to be brought to the state’s TAC for iterative discussion and feedback at both the design stage and when results are available.

Expanding the Linking item set. In both fixed form and adaptive testing, some creativity may be required. There are essentially three potential sets of intersections of the dimensions of the NGSS that can be included on assessments. They include those that are addressed by the performance expectations, the expanded set addressed by the topics³⁹, and the complete set of intersections in the total NGSS domain. Because nearly every state has moved toward online assessment, there is greater flexibility to expand the linking item set without the substantial cost of developing and printing many unique forms.

One such approach is depicted in Figure 11. The largest of the concentric circles represents the bounds of the complete NGSS domain (e.g., it contains approximately the same number of dots as there are intersections in the middle school NGSS). The middle-sized circle represents the intersections of the NGSS addressed by the collection of topics in the middle school NGSS. The smallest circle represents the intersections of the NGSS addressed by the collection of expectations in the middle school NGSS.

Figure 11. A hypothetical approach to expanding the linking item set.



In this hypothetical scenario, the state has developed item sets that address approximately one third of the intersections in the expectations (the empty light blue dots), and approximately ten percent of the intersections in the topics (the empty light red dots). This state has not developed any item sets addressing intersections outside the

³⁹ See footnote 6: some states are expanding on the very limited set of intersections in the *expectations* by developing test content based on the complete set of interactions in the *topics*, even though most of those interactions do not appear in any *performance expectations*.

topics. This state has decided to address the issue of transfer using a core and matrix design. In this design, the item sets measuring the intersections of the expectations are treated as the core, which are eligible to appear on all test forms in any given year. Those that serve as the core in year 1 are shown with filled black dots. Those that serve as the core in year 2 are shown with filled gray dots.

The item sets measuring the intersection of the topics outside the expectations are treated as the matrix, meaning that these item sets are eligible to appear on any given year's forms. However, which item sets appear change each year to address the issue of transfer. Those that appear on year-1 forms are shown with filled red dots. Those that appear in year two are shown with filled blue dots, and those that appear in year 3 are shown with filled yellow dots.

To increase the size of the linking item set and the quality of equating, item sets that appear as core item sets in one year are always matrixed the next year so that the entire core of item sets can be used for equating the next year. In this manner, every student takes each of the sets of items described in Table 5 in a given year where each item set is described in detail.

Table 5. A core and matrix design intended to balance focus, transfer, and the need for strong within- and across-year equating.

Item Set Description	Purpose			Based on...		Number	
	Equating		Balance Focus & Transfer	Expectations	Topics	Available	Administered per student
	Within- Year	Across- year					
Core item set (will be <i>matrixed</i> on next year's test)	✓	-	-	✓	-	8	8
Matrixed item set 1 (from last year's <i>core</i> item set)	-	✓	-	✓	-	8	1
Matrixed item set 2 (not used on next year's test)	-	-	✓	-	✓	8	1
				Total		24	10

References

- Alabama State Department of Education. (2016). *Alabama Benchmark Scores: Achievement Proficiency for Public Reporting. ACT Aspire 2015-2016*. Montgomery, AL: Author. Retrieved August 30, 2017, from https://www.alsde.edu/sec/sa/Testing/Alabama_Proficiency_Levels_Table_2015-2016.pdf
- Andrich, D. (1983). Recognizing a Violation of the Assumption of Local Independence in the Analysis of Test Data. *Educational and Psychological Measurement, 43*(3), 829-833.
- Bolt, D. M. (1999). Evaluating the Effects of Multidimensionality on IRT True-Score Equating. *Applied Measurement in Education, 12*(4), 383-407.
- Bolt, D. M., & Kim, J.-S. (2015). Hierarchical Item Response Theory Modeling. In *Encyclopedia of Statistics in Behavioral Science*.
- Brossman, B. G., & Lee, W.-C. (2013). Observed Score and True Score Equating Procedures for Multidimensional Item Response Theory. *Applied Psychological Measurement, 37*(6), 460-481.
- Camilli, G., Wang, M.-M., & Fesq, J. (1995). The Effects of Dimensionality on Equating the Law School Admission Test. *Journal of Educational Measurement, 32*(1), 79-96.
- Cao, Y. (2008). *Mixed-Format Test Equating: Effects of Test Dimensionality and Common-Item Sets*. Unpublished Dissertation, University of Maryland.
- Cao, Y., Lu, R., & Tao, W. (2014). Effect of Item Response Theory (IRT) Model Selection on Testlet-Based Test Equating. *ETS Research Report Series, 2014*(2). Retrieved September 13, 2017, from <http://onlinelibrary.wiley.com/doi/10.1002/ets2.12017/pdf>
- Chen, J. (2014). *Model Selection for IRT Equating of Testlet-Based Tests in the Random Groups Design*. Unpublished dissertation, University of Iowa.
- Cohen, J. (2017). *IRT Model for Item Clusters*. Portsmouth, NH: Presentation at the annual Reidy Interactive Lecture Series of the Center for Assessment.
- Committee on a Conceptual Framework for New K-12 Science Education Standards. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press.
- Davey, T., Oshima, T., & Lee, K. (1996). Linking Multidimensional Item Calibrations. *Applied Psychological Measurement, 20*(4), 405-416.
- De Champlain, A. F. (1996). The Effect of Multidimensionality on IRT True-Score Equating for Subgroups of Examinees. *Journal of Educational Measurement, 33*(2), 181-201.
- Doran, H. (2017). *NGSS IRT Model*. Washington, DC: American Institutes for Research. (working paper, not available for distribution).
- Dorans, N. J., & Kingston, N. M. (1985). The Effects of Violations of Unidimensionality on the Estimation of Item and Ability Parameters and Item Response Theory Equating of the GRE Verbal Scale. *Journal of Educational Measurement, 22*(4), 249-262.
- Hirsch, T. (1989). Multidimensional Equating. *Journal of Educational Measurement, 26*(4), 337-349.
- Jiao, H., & Lissitz, R. (2015). Modeling Latent Growth Using Mixture Item Response Theory. In R. Lissitz, & H. Jiao (Eds.), *Value Added Modeling and Growth Modeling with Particular Application to Teacher and School Effectiveness* (pp. 293-306). Charlotte, NC: Information Age Publishing, Inc.
- Jiao, H., & Zhang, U. (2017). Polytomous Multilevel Testlet Models for Testlet-based Assessments with Complex Sampling Designs. *British Journal of Mathematical and Statistical Psychology, 68*(1), 65-83.
- Li, Y., Bolt, D. M., & Fu, J. (2006). A Comparison of Alternative Models for Testlets. *Applied Psychological Measurement, 30*(1), 3-21.
- Li, Y., Jiao, H., & Lissitz, R. W. (2012). Applying Multidimensional Item Response Theory Models in Validating Test Dimensionality: An Example of K-12 Large-scale Science Assessments. *Journal of Applied Testing Technology, 13*(2). Retrieved September 13, 2017, from <http://www.testpublishers.org/assets/applying%20multidimensional%20item.pdf>
- Louisiana Department of Education. (n.d.). *Grade 3-8 Assessments Fact Sheet*. Author. Retrieved September 12, 2017, from <http://www.louisianabelieves.com/docs/default-source/assessment/grades-3-8-assessments-fact-sheet.pdf?sfvrsn=2>
- Martineau, J. A., & Zeng, J. (2015). Borrowing the Strength of Unidimensional Scaling to Produce Multidimensional Educational Effectiveness Profiles. In R. W. Lissitz, & H. Jiao (Eds.), *Value Added Modeling and Growth Modeling with Particular Application to Teacher and School Effectiveness* (pp. 103-138). Charlotte, NC: Information Age Publishing.
- Mislevy, R. J., Levy, R., Kroopnick, M., & Rutstein, D. (2007). *Evidentiary Foundations of Mixture Item Response Theory Models*. Retrieved September 13, 2017, from <http://www.education.umd.edu/EDMS/mislevy/papers/Mixtures.doc>
- Monaghan, W. (2006, July). The Facts About Subscores. *R & D Connections*. Retrieved September 13, 2017, from https://www.ets.org/Media/Research/pdf/RD_Connections4.pdf

- Nandakumar, R. (1991). Traditional Dimensionality Versus Essential Dimensionality. *Journal of Educational Measurement*, 28(2), 99-117. Retrieved September 13, 2017, from https://www.jstor.org/stable/1434793?seq=1#page_scan_tab_contents
- NGSS Lead States. (2013a). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press. Retrieved August 31, 2017, from <https://www.nextgenscience.org>
- NGSS Lead States. (2013b). *Next Generation Science Standards: For States, By States - Three Dimensional Learning*. Washington, DC: The National Academies Press. Retrieved August 27, 2017, from <https://www.nextgenscience.org/three-dimensions>
- NGSS Lead States. (2013c). *Next Generation Science Standards: For States, By States - DCI Arrangements of the Next Generation Science Standards*. Washington, DC: The National Academies Press. Retrieved August 31, 2017, from <http://www.nextgenscience.org/sites/default/files/NGSS%20DCI%20Combined%2011.6.13.pdf>
- NGSS Lead States. (2013d). *Next Generation Science Standards: For States, By States - Appendix J - Science, Technology, Society and the Environment*. Washington, DC: The National Academies Press. Retrieved August 31, 2017, from https://www.nextgenscience.org/sites/default/files/APPENDIX%20J_0.pdf
- NGSS Lead States. (2013e). *Next Generation Science Standards: For States, By States - Appendix H - Understanding the Scientific Enterprise: The Nature of Science in the Next Generation Science Standards*. Washington, DC: The National Academies Press. Retrieved August 31, 2017, from <https://www.nextgenscience.org/sites/default/files/Appendix%20H%20-%20The%20Nature%20of%20Science%20in%20the%20Next%20Generation%20Science%20Standards%204.15.13.pdf>
- NGSS Lead States. (2013f). *Next Generation Science Standards: For States, By States - NGSS [High School] Evidence Statements*. Washington, DC: National Academies Press.
- NGSS Lead States. (2013g). *Next Generation Science Standards: For States, By States - NGSS [Elementary School] Evidence Statements*. Washington, DC: National Academies Press.
- NGSS Lead States. (2013h). *Next Generation Science Standards: For States, By States - NGSS [Middle School] Evidence Statements*. Washington, DC: National Academies Press.
- Pellegrino, J. (2017). *Measuring What Matters: Opportunities & Challenges in the Equitable Assessment of Science Proficiency*. Presentation at the June 2017 meeting of the Technical Issues in Large Scale Assessment (TILSA) State Collaborative on Assessment and Student Standards (SCASS) of the Council of Chief State School Officers (CCSSO). Austin, TX.
- Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, 47(5), 667-696.
- Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and Structural Coefficient Bias in Structural Equation Modeling: A Bifactor Perspective. *Educational and Psychological Measurement*, 73(1), 5-26. doi:10.1177/0013164412449831
- Rijmen, F. (2017). *An IRT Model for NGSS*. Washington, DC: American Institutes for Research. (working paper, not available for distribution).
- Sideridis, G. D. (2011). The Effects of Local Item Dependence on Estimates of Ability in the Rasch Model. *Rasch Measurement Transactions*, 25(3), 1334-1336.
- Utah State Board of Education. (2017). *Assessments*. Salt Lake City, UT: Author. Retrieved August 30, 2017, from <https://www.schools.utah.gov/assessment/assessments>
- Zeng, J. (2010). *Development of a Hybrid Method for Dimensionality Identification Incorporating an Angle-Based Approach*. Unpublished dissertation. Ann Arbor, MI: University of Michigan. Retrieved September 13, 2017, from https://deepblue.lib.umich.edu/bitstream/handle/2027.42/75874/jzeng_1.pdf?sequence=1&isAllowed=y
- Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2002). Identification and Evaluation of Local Item Dependencies in the Medical College Admissions Test. *Journal of Educational Measurement*, 39(4), 291-309.

APPENDIX A: NGSS Domain Coverage by Grade or Grade Span

The domain coverage of the expectations at the various grades and grade levels is provided in Table A.1. I added grade spans K-2, 3-12, and K-12 even though they are unlikely to form the basis of any NGSS assessment. These are interesting just because they provide an understanding of domain coverage in lower grades and in broader grade spans. Domain coverage for each of these grades and grade spans is displayed graphically in the figures on the remaining pages.

Of note, in evaluating core ideas, practices, concepts, and nature of science elements *by themselves*, there are some parts of the domain that are not address by *any* expectation in *any* grade, across tested grades (3-12) or all grades (K-12). Two percent of core ideas and twelve percent of concepts are not addressed by any expectation.

In evaluating the combinations (or the complete set of cells in the matrix), there are also potential concerns. Less than twelve percent of cells in the domain are covered by *any* expectation in *any* grade.

Table A.1. Domain coverage by expectations from each grade or grade span.

Dimension or Component	Type	Code	Percent of Elements of a Dimension or Component Addressed in One or More Expectations or in One or More Topics											
			K	1	2	3	4	5	K-2	3-5	MS	HS	3-12	K-12
Disciplinary Core Ideas	Dimension	DCI	33	29	27	38	44	31	71	91	89	93	96	98
Science & Engineering Practices	Dimension	SEP	88	75	88	88	88	75	88	100	100	100	100	100
Crosscutting Concepts	Dimension	CCC	57	43	71	57	57	71	86	71	100	100	100	100
Nature of Science	Component	NOS	25	38	38	50	38	38	63	75	75	88	88	88
Combination of Dimensions and/or Components		Part of Domain	Percent of Domain Cells Addressed in Topics											
			K	1	2	3	4	5	K-2	3-5	MS	HS	3-12	K-12
DCI × SEP × CCC		3D	2.9	1.7	3.2	3.2	2.6	1.6	6.3	6.6	24.4	28.3	43.3	45.2
(DCI × SEP × CCC) + (DCI × NOS)		3D + 1	2.8	1.8	3.1	3.3	2.7	1.6	6.5	6.8	24.1	27.9	43.0	45.0
Combination of Dimensions and/or Components		Part of Domain	Percent of Domain Cells Addressed in Expectations											
			K	1	2	3	4	5	K-2	3-5	MS	HS	3-12	K-12
DCI × SEP × CCC		3D	0.9	0.5	0.6	0.8	1.0	0.6	1.7	2.3	3.2	3.8	8.0	9.3
(DCI × SEP × CCC) + (DCI × NOS)		3D + 1	1.0	0.5	0.7	0.9	1.1	0.7	2.0	2.5	3.7	4.5	9.3	10.6

