The Intersection of Measurement Model, Equating, and the Next Generation Science Standards

Joseph A. Martineau
Senior Associate

High School Domain Coverage

	SEP 1: Ask questions							
	CCC 0: none	CCC 1: patterns	CCC 2: cause & effect	CCC 3: scale & quantity	CCC 4: systems/models	CCC 5: energy & matter	CCC 6: structure & function	CCC 7: stability & change
DCI 01: PS1.A								
DCI 02: PS1.B								
DCI 03: PS1.C								
DCI 04: PS2.1								
DCI 05: PS2.B								
DCI 06: PS2.C								
DCI 07: PS3.A								
DCI 08: PS3.B								
DCI 09: PS3.C								
DCI 10: PS3.D								
DCI 11: PS3.E								
DCI 12: PS4.A								1
DCI 13: PS4.B								
DCI 14: PS4.C								
DCI 15: LS1.A			1					
DCI 16: LS1.B								
DCI 17: LS1.C								
DCI 18: LS1.D								
DCI 19: LS2.A								
DCI 20: LS2.B								
DCI 21: LS2.C								
DCI 22: LS2.D								
DCI 23: LS3.A			1					
DCI 24: LS3.B								
DCI 25: LS4.A								
DCI 26: LS4.B								
DCI 27: LS4.C								
DCI 28: LS4.D								
DCI 29: ESS1.A								
DCI 30: ESS1.B								
DCI 31: ESS1.C								
DCI 32: ESS2.A								
DCI 33: ESS2.B								
DCI 34: ESS2.C								
DCI 35: ESS2.D								
DCI 36: ESS2.E								
DCI 37: ESS3.A								
DCI 38: ESS3.B								
DCI 39: ESS3.C								
DCI 40: ESS3.D								
DCI 41: ETS1.A								
DCI 42: ETS1.B								
DCI 43: ETS1.C								
DCI 44: ETS2.A								
DCI 45: ETS2.B								1

	SEP 2: Modeling							
	CCC 0: none	CCC 1: patterns	CCC 2: cause & effect	CCC 3: scale & quantity	CCC 4: systems/models	CCC 5: energy & matter	CCC 6: structure & function	CCC 7: stability & change
DCI 01: PS1.A		1				1		
DCI 02: PS1.B						1		
DCI 03: PS1.C						1		
DCI 04: PS2.1								
DCI 05: PS2.B								
DCI 06: PS2.C								
DCI 07: PS3.A						1		
DCI 08: PS3.B								
DCI 09: PS3.C			1					
DCI 10: PS3.D				1	1			
DCI 11: PS3.E								
DCI 12: PS4.A								
DCI 13: PS4.B								
DCI 14: PS4.C								
DCI 15: LS1.A					1			
DCI 16: LS1.B					1			
DCI 17: LS1.C						2		
DCI 18: LS1.D								
DCI 19: LS2.A								
DCI 20: LS2.B					1			
DCI 21: LS2.C								
DCI 22: LS2.D								
DCI 23: LS3.A								
DCI 24: LS3.B								
DCI 25: LS4.A								
DCI 26: LS4.B								
DCI 27: LS4.C								
DCI 28: LS4.D								
DCI 29: ESS1.A				1				
DCI 30: ESS1.B		1						
DCI 31: ESS1.C								
DCI 32: ESS2.A		1				1		1
DCI 33: ESS2.B						1		1
DCI 34: ESS2.C								
DCI 35: ESS2.D		1				1		
DCI 36: ESS2.E								
DCI 37: ESS3.A								
DCI 38: ESS3.B								
DCI 39: ESS3.C								
DCI 40: ESS3.D								
DCI 41: ETS1.A								
DCI 42: ETS1.B								
DCI 43: ETS1.C								
DCI 44: ETS2.A						1		
DCI 45: ETS2.B								

	SEP 3: Investigation							
	CCC 0: none	CCC 1: patterns	CCC 2: cause & effect	CCC 3: scale & quantity	CCC 4: systems/models	CCC 5: energy & matter	CCC 6: structure & function	CCC 7: stability & change
DCI 01: PS1.A		1						
DCI 02: PS1.B								
DCI 03: PS1.C								
DCI 04: PS2.1								
DCI 05: PS2.B			1					
DCI 06: PS2.C								
DCI 07: PS3.A			1					
DCI 08: PS3.B					1			
DCI 09: PS3.C								
DCI 10: PS3.D					1			
DCI 11: PS3.E								
DCI 12: PS4.A								
DCI 13: PS4.B								
DCI 14: PS4.C								
DCI 15: LS1.A								1
DCI 16: LS1.B								
DCI 17: LS1.C								
DCI 18: LS1.D								
DCI 19: LS2.A								
DCI 20: LS2.B								
DCI 21: LS2.C								
DCI 22: LS2.D								
DCI 23: LS3.A								
DCI 24: LS3.B				1				
DCI 25: LS4.A								
DCI 26: LS4.B								
DCI 27: LS4.C								
DCI 28: LS4.D								
DCI 29: ESS1.A								
DCI 30: ESS1.B								
DCI 31: ESS1.C								
DCI 32: ESS2.A								1
DCI 33: ESS2.B								
DCI 34: ESS2.C							1	
DCI 35: ESS2.D								1
DCI 36: ESS2.E								
DCI 37: ESS3.A								
DCI 38: ESS3.B								
DCI 39: ESS3.C								
DCI 40: ESS3.D								1
DCI 41: ETS1.A								
DCI 42: ETS1.B								
DCI 43: ETS1.C								
DCI 44: ETS2.A								
DCI 45: ETS2.B								1

	SEP 4: Using data							
	CCC 0: none	CCC 1: patterns	CCC 2: cause & effect	CCC 3: scale & quantity	CCC 4: systems/models	CCC 5: energy & matter	CCC 6: structure & function	CCC 7: stability & change
DCI 01: PS1.A								
DCI 02: PS1.B								
DCI 03: PS1.C								
DCI 04: PS2.1			1					
DCI 05: PS2.B								
DCI 06: PS2.C								
DCI 07: PS3.A								
DCI 08: PS3.B								
DCI 09: PS3.C								
DCI 10: PS3.D								
DCI 11: PS3.E								
DCI 12: PS4.A								
DCI 13: PS4.B								
DCI 14: PS4.C								
DCI 15: LS1.A								
DCI 16: LS1.B								
DCI 17: LS1.C								
DCI 18: LS1.D								
DCI 19: LS2.A								
DCI 20: LS2.B								
DCI 21: LS2.C								
DCI 22: LS2.D								
DCI 23: LS3.A								
DCI 24: LS3.B				1				
DCI 25: LS4.A								
DCI 26: LS4.B		1						
DCI 27: LS4.C		1						
DCI 28: LS4.D								
DCI 29: ESS1.A								
DCI 30: ESS1.B								
DCI 31: ESS1.C								
DCI 32: ESS2.A								
DCI 33: ESS2.B								
DCI 34: ESS2.C								
DCI 35: ESS2.D								
DCI 36: ESS2.E								
DCI 37: ESS3.A								
DCI 38: ESS3.B								
DCI 39: ESS3.C								
DCI 40: ESS3.D								
DCI 41: ETS1.A								
DCI 42: ETS1.B								
DCI 43: ETS1.C								
DCI 44: ETS2.A								
DCI 45: ETS2.B								

	SEP 5: Mathematical thinking							
	CCC 0: none	CCC 1: patterns	CCC 2: cause & effect	CCC 3: scale & quantity	CCC 4: systems/models	CCC 5: energy & matter	CCC 6: structure & function	CCC 7: stability & change
DCI 01: PS1.A								
DCI 02: PS1.B						1		
DCI 03: PS1.C								
DCI 04: PS2.1					1			
DCI 05: PS2.B		1						
DCI 06: PS2.C								
DCI 07: PS3.A					1			
DCI 08: PS3.B					1			
DCI 09: PS3.C								
DCI 10: PS3.D								
DCI 11: PS3.E								
DCI 12: PS4.A			1					
DCI 13: PS4.B								
DCI 14: PS4.C								
DCI 15: LS1.A								
DCI 16: LS1.B								
DCI 17: LS1.C								
DCI 18: LS1.D								
DCI 19: LS2.A				2				
DCI 20: LS2.B						1		
DCI 21: LS2.C				1				
DCI 22: LS2.D								
DCI 23: LS3.A								
DCI 24: LS3.B								
DCI 25: LS4.A								
DCI 26: LS4.B								
DCI 27: LS4.C			1					
DCI 28: LS4.D			1					
DCI 29: ESS1.A								
DCI 30: ESS1.B				1				
DCI 31: ESS1.C								
DCI 32: ESS2.A								
DCI 33: ESS2.B								
DCI 34: ESS2.C								
DCI 35: ESS2.D					1			
DCI 36: ESS2.E								
DCI 37: ESS3.A								
DCI 38: ESS3.B								
DCI 39: ESS3.C								1
DCI 40: ESS3.D					1			
DCI 41: ETS1.A								
DCI 42: ETS1.B		1		1				
DCI 43: ETS1.C								
DCI 44: ETS2.A				1				
DCI 45: ETS2.B								1

	SEP 6: Construct explanations							
	CCC 0: none	CCC 1: patterns	CCC 2: cause & effect	CCC 3: scale & quantity	CCC 4: systems/models	CCC 5: energy & matter	CCC 6: structure & function	CCC 7: stability & change
DCI 01: PS1.A		1						
DCI 02: PS1.B		2						1
DCI 03: PS1.C								1
DCI 04: PS2.1			1					
DCI 05: PS2.B								
DCI 06: PS2.C								
DCI 07: PS3.A						1		
DCI 08: PS3.B								
DCI 09: PS3.C								
DCI 10: PS3.D						1		
DCI 11: PS3.E								
DCI 12: PS4.A					1			
DCI 13: PS4.B					1			
DCI 14: PS4.C								
DCI 15: LS1.A							1	
DCI 16: LS1.B								
DCI 17: LS1.C						1		
DCI 18: LS1.D								
DCI 19: LS2.A								
DCI 20: LS2.B						1		
DCI 21: LS2.C							1	
DCI 22: LS2.D								
DCI 23: LS3.A								
DCI 24: LS3.B								
DCI 25: LS4.A								
DCI 26: LS4.B			1					
DCI 27: LS4.C			2					
DCI 28: LS4.D								1
DCI 29: ESS1.A						1		
DCI 30: ESS1.B								
DCI 31: ESS1.C								1
DCI 32: ESS2.A								
DCI 33: ESS2.B								
DCI 34: ESS2.C								
DCI 35: ESS2.D								
DCI 36: ESS2.E								
DCI 37: ESS3.A			1					
DCI 38: ESS3.B			1					
DCI 39: ESS3.C								1
DCI 40: ESS3.D								
DCI 41: ETS1.A			1			1		
DCI 42: ETS1.B								2
DCI 43: ETS1.C		1	1					1
DCI 44: ETS2.A						1		
DCI 45: ETS2.B			1			1		1

	SEP 7: Argue from evidence							
	CCC 0: none	CCC 1: patterns	CCC 2: cause & effect	CCC 3: scale & quantity	CCC 4: systems/models	CCC 5: energy & matter	CCC 6: structure & function	CCC 7: stability & change
DCI 01: PS1.A								
DCI 02: PS1.B								
DCI 03: PS1.C		1						
DCI 04: PS2.1								
DCI 05: PS2.B								
DCI 06: PS2.C								
DCI 07: PS3.A								
DCI 08: PS3.B								
DCI 09: PS3.C								
DCI 10: PS3.D								
DCI 11: PS3.E								
DCI 12: PS4.A					1			
DCI 13: PS4.B					1			
DCI 14: PS4.C								
DCI 15: LS1.A								
DCI 16: LS1.B								
DCI 17: LS1.C								
DCI 18: LS1.D								
DCI 19: LS2.A								
DCI 20: LS2.B								
DCI 21: LS2.C								1
DCI 22: LS2.D			1					
DCI 23: LS3.A								
DCI 24: LS3.B			1					
DCI 25: LS4.A								
DCI 26: LS4.B								
DCI 27: LS4.C			1					
DCI 28: LS4.D								
DCI 29: ESS1.A								
DCI 30: ESS1.B								
DCI 31: ESS1.C		1						
DCI 32: ESS2.A								
DCI 33: ESS2.B		1						
DCI 34: ESS2.C								
DCI 35: ESS2.D								1
DCI 36: ESS2.E								1
DCI 37: ESS3.A								
DCI 38: ESS3.B								
DCI 39: ESS3.C								
DCI 40: ESS3.D								
DCI 41: ETS1.A								
DCI 42: ETS1.B								
DCI 43: ETS1.C								
DCI 44: ETS2.A								
DCI 45: ETS2.B								

	SEP 8: Using information							
	CCC 0: none	CCC 1: patterns	CCC 2: cause & effect	CCC 3: scale & quantity	CCC 4: systems/models	CCC 5: energy & matter	CCC 6: structure & function	CCC 7: stability & change
DCI 01: PS1.A								
DCI 02: PS1.B								
DCI 03: PS1.C								
DCI 04: PS2.1								
DCI 05: PS2.B							1	
DCI 06: PS2.C								
DCI 07: PS3.A								
DCI 08: PS3.B								
DCI 09: PS3.C								
DCI 10: PS3.D			1					
DCI 11: PS3.E								
DCI 12: PS4.A			1					
DCI 13: PS4.B			2					
DCI 14: PS4.C			1					
DCI 15: LS1.A								
DCI 16: LS1.B								
DCI 17: LS1.C								
DCI 18: LS1.D								
DCI 19: LS2.A								
DCI 20: LS2.B								
DCI 21: LS2.C								
DCI 22: LS2.D								
DCI 23: LS3.A								
DCI 24: LS3.B								
DCI 25: LS4.A		1						
DCI 26: LS4.B								
DCI 27: LS4.C								
DCI 28: LS4.D								
DCI 29: ESS1.A						1		
DCI 30: ESS1.B								
DCI 31: ESS1.C								
DCI 32: ESS2.A								
DCI 33: ESS2.B								
DCI 34: ESS2.C								
DCI 35: ESS2.D								
DCI 36: ESS2.E								
DCI 37: ESS3.A								
DCI 38: ESS3.B								
DCI 39: ESS3.C								
DCI 40: ESS3.D								
DCI 41: ETS1.A								
DCI 42: ETS1.B								
DCI 43: ETS1.C								
DCI 44: ETS2.A			1					
DCI 45: ETS2.B			1					

	NOS 1: variety of methods	NOS 2: based on evidence	NOS 3: open to revision	NOS 4: explains phenomena	NOS 5: way of knowing	NOS 6: order & consistency	NOS 7: human endeavor	NOS 8: addresses questions
DCI 01: PS1.A								
DCI 02: PS1.B						1		
DCI 03: PS1.C				1				
DCI 04: PS2.1				1				
DCI 05: PS2.B				1				
DCI 06: PS2.C								
DCI 07: PS3.A						1		
DCI 08: PS3.B						1		
DCI 09: PS3.C								
DCI 10: PS3.D								
DCI 11: PS3.E								
DCI 12: PS4.A				1				
DCI 13: PS4.B				2		1		
DCI 14: PS4.C								
DCI 15: LS1.A	1							
DCI 16: LS1.B								
DCI 17: LS1.C								
DCI 18: LS1.D								
DCI 19: LS2.A			1					
DCI 20: LS2.B			1					
DCI 21: LS2.C			2					
DCI 22: LS2.D			1					
DCI 23: LS3.A								
DCI 24: LS3.B							1	
DCI 25: LS4.A				1		1		
DCI 26: LS4.B								
DCI 27: LS4.C						1		
DCI 28: LS4.D								
DCI 29: ESS1.A			1			1		
DCI 30: ESS1.B		1						
DCI 31: ESS1.C				1				
DCI 32: ESS2.A		2						
DCI 33: ESS2.B		1						
DCI 34: ESS2.C								
DCI 35: ESS2.D		1						
DCI 36: ESS2.E								
DCI 37: ESS3.A								1
DCI 38: ESS3.B								
DCI 39: ESS3.C							1	
DCI 40: ESS3.D	1	1						
DCI 41: ETS1.A								
DCI 42: ETS1.B								1
DCI 43: ETS1.C								
DCI 44: ETS2.A		1		1		1		
DCI 45: ETS2.B							1	1

Implication of Domain Structure: Sparse Domain Coverage

Dimension or Component	Type	Code	Element Coverage		
			3-5	MS	HS
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**	***Cell Coverage***		
			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%

- *Expectations* are focused on a small number of intersections of the domain to make the infeasibly large complete domain manageable in the K-12 context.
- Leads to the sparse domain coverage
- A key practical implication of sparse domain coverage

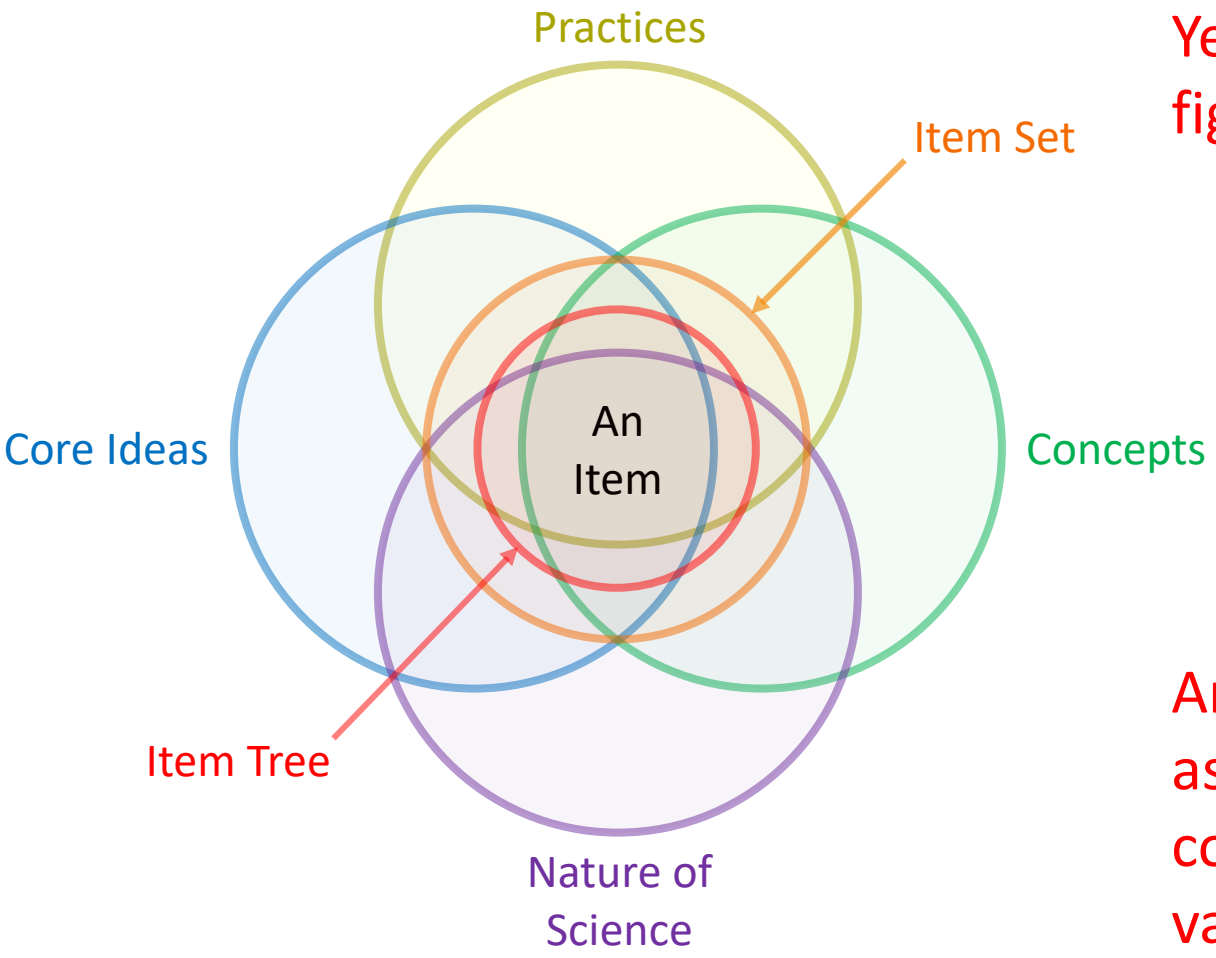
Do we value focus? It's OK to *de facto* define K-12 science as the specific intersections of the elements of each dimension of the domain covered by existing *expectations* (and assume that transfer happens)

Do we value transfer? Novel *expectations* are necessary at least to some degree to address transfer

Additional Potential Implications of the Domain Structure

- Nested data structures, with cross-nesting and multiple levels
  - Item sets
    - Groups of items are attached to the same set of [potentially interactive] stimulus material
  - Item trees
    - Items administered on given a certain response on a previous item
    - Items scorable only in light of the response on another item
  - Scoring assertions
    - An AIR term
    - There may be multiple scores per item, or across items, depending on the interactions of the student with the items
  - Cross-nesting within dimensions of the NGSS

Nested Data Structures – An Ambiguously Useful Depiction



Yeah. Not so helpful as an interpretable figure.

And it doesn't even incorporate scoring assertions. But it is instructive of the complexity of the complete set of nesting variables that may need to be accounted for.

Multidimensionality!

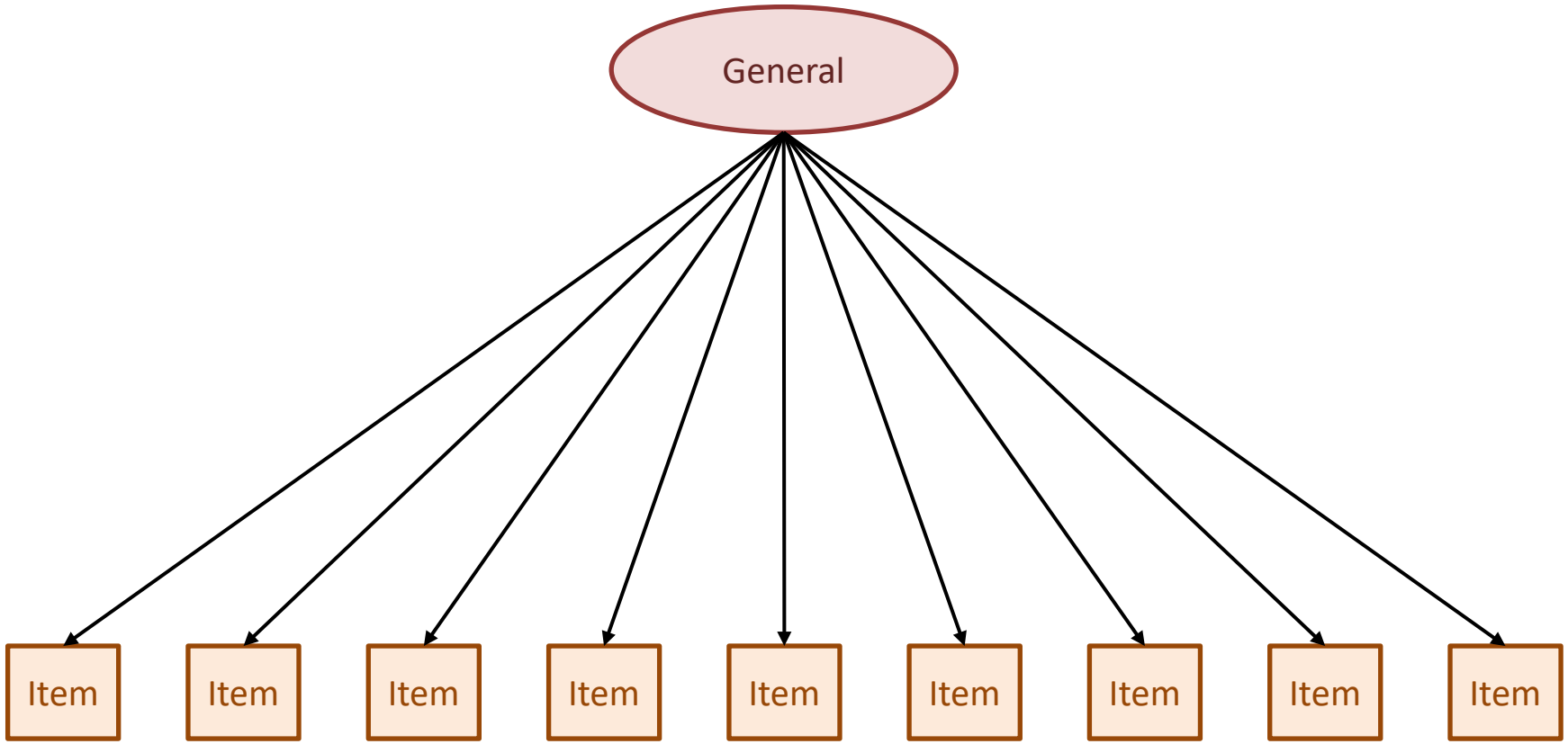
We need to stop trying to find subscores. We should be constructing them from the beginning.

—Nathan Dadey (late Wednesday night)

Additional Potential Implications of the Domain Structure

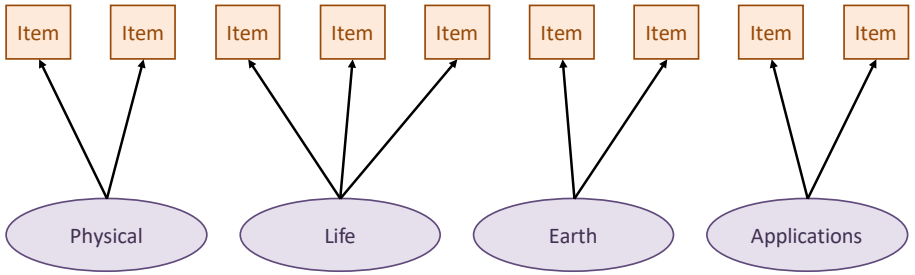
- The NGSS is structured differently than other standards we have seen.
- Regardless of care to *construct* subscores, many components of the NGSS could turn out to be psychometric dimensions (a different definition than is used in the NGSS)
  - The four disciplines
  - Clusters of the science and engineering practices
  - Clusters of the cross-cutting concepts
  - Clusters of the nature of science
  - Item type
  - Item sets
- The data could also be essentially unidimensional
- Score reports and empirical dimensionality do not need to be mirrors of each other. In fact there may be good reasons for them to be different.
  - We often report subscores an a unidimensionally-modeled construct.
  - We can also report fewer subscores on a multidimensionally-modeled construct.

Dimensionality – Maybe Just This (Yay!)

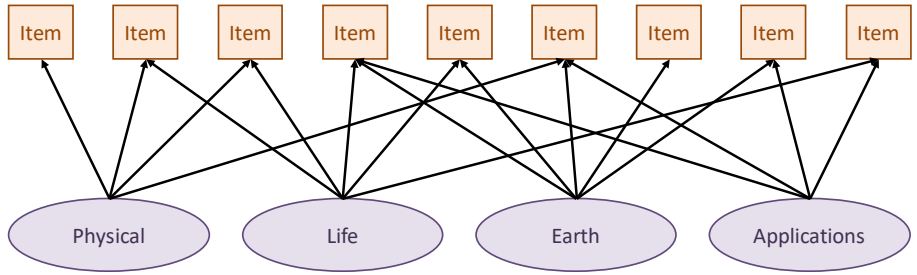


Dimensionality – Potential Scenarios

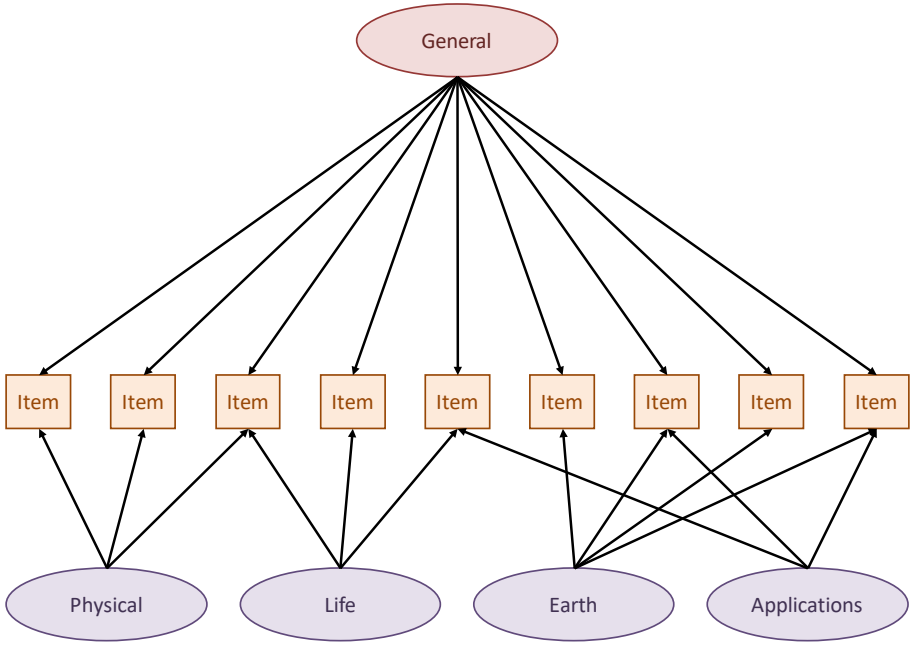
Simple Structure



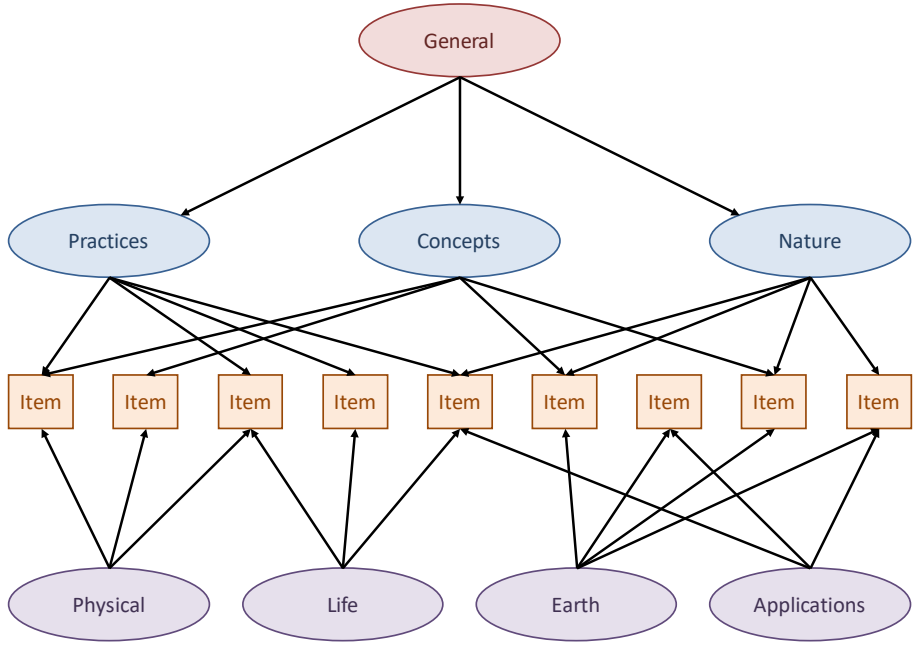
Complex Structure



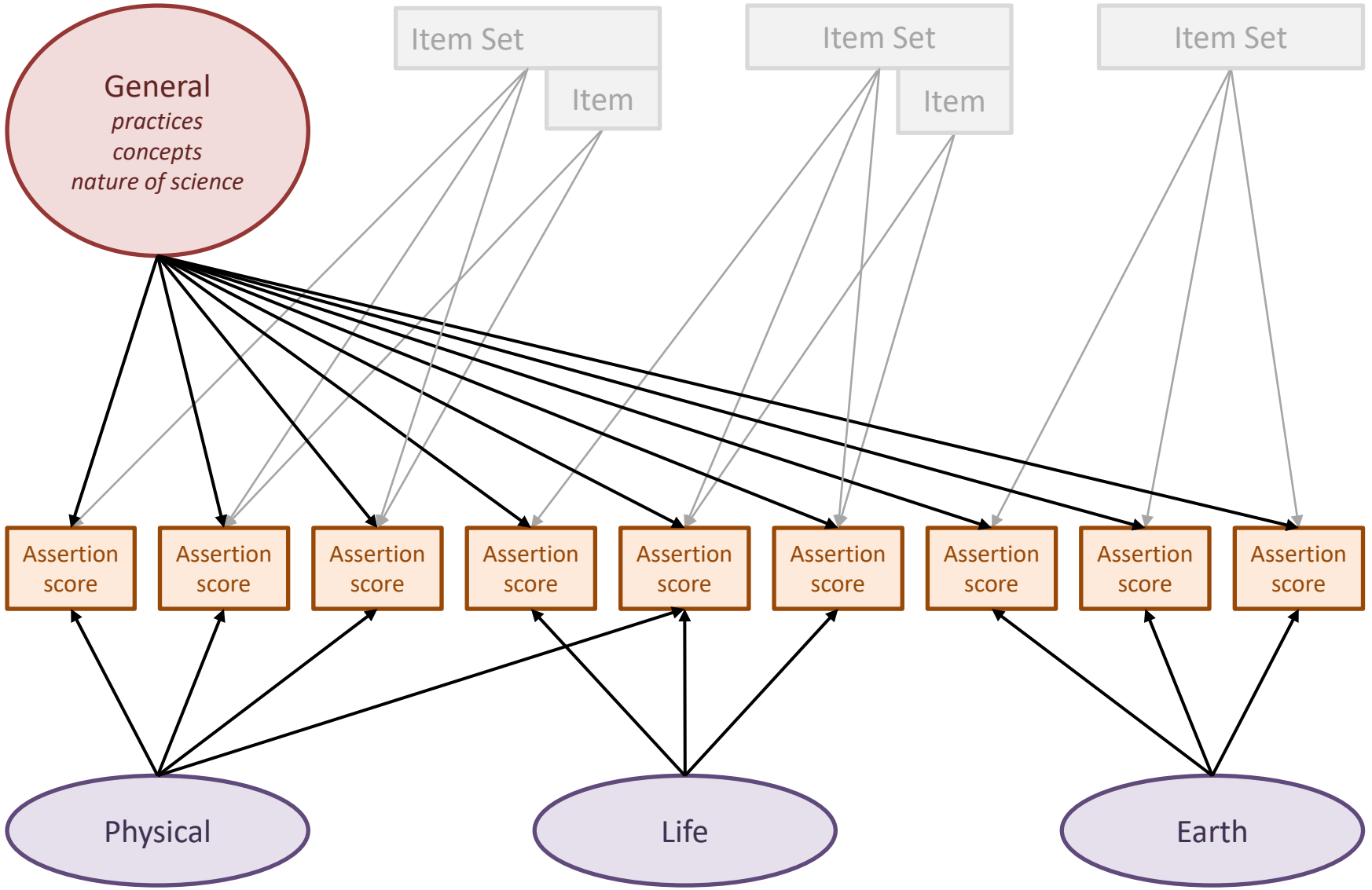
Bifactor



Higher Order



Dimensionality – My Somewhat Educated Guess



Implications for Selecting a Measurement Model

- Nested data structures and dimensionality will theoretically will be reflected in relationships among item and/or score variables.
  - Local item dependence can come from either unmodeled nested data structures or unmodeled dimensionality.
  - But, we have long experience with item and score data behaving in ways we don't expect.
  - An analogy: evolution of a psychometrician's model of his toddler's utterances…

- It may be that a simple unidimensional model works, or a more complex model may be needed to reflect nesting and multidimensionality.

Implications for Selecting a Measurement Model

- If a more complex model is needed, it is likely to be unfamiliar territory for state assessment staff and most vendor staff.
  - Enlist your vendor (and in-house psychometrician, if available) to design measurement model selection studies
    - If the field test strongly resembles the operational test, it is OK to use field test data.
    - If not, it is better to wait for operational data, even if it means an additional delay in reporting.
  - Enlist your TAC to review and provide feedback on:
    - The design of measurement model selection studies.
    - Results of model selection studies and associated recommendations.
    - The completeness of documentation of the implementation to allow for transition between vendors.
  - Re-conduct the studies in later cycles to evaluate ongoing fit of the measurement model
    - Don’t assume that early year data will resemble later data.
    - Your TAC can recommend appropriate intervals

Implications for Equating

- Novel or more complex models
  - Equating will also be more complex, in the form of either…
    - Using specialized equating methods developed for the measurement model, or
    - Making some assumptions and equating on true score or raw score…
  - Greater size of the linking item set needed…
  - AND issues associated with non-representativeness in using item clusters in anchor item sets
    - With online assessment, states and vendors have more flexibility to make this feasible
    - Last year's entire item pool can be used to equate by matrix-sampling the anchor item set by student
    - Less costly than lots of paper forms
  - Thank you, Jon Cohen for this D'OH! moment (for me at least) this morning at breakfast
  - May also be new territory for state staff and vendor staff
  - Recommendations for TAC review and feedback involvement:
    - Thorough documentation of equating design, with justification for why the design decisions were appropriate
    - Equating results
    - Scale stability study design
    - Scale stability study results
    - Thorough documentation of equating business rules

An Inherent Tension: Focus vs. Transfer

- Important to explicitly deliberate on the tension issue
  - There are valid rationales for valuing balance of focus and transfer along the entire continuum

- Privileging focus means existing expectations are all that will be developed
  - The domain of science as described in the PEs *is* K-12 science
  - Assumes that transfer will occur if the PEs are well taught
  - This decision introduces no additional complexities to equating

An Inherent Tension: Focus vs. Transfer

- Privileging transfer means that content is built out over time to more reasonably cover the matrix
  - Every expectation for which items have been developed is eligible for inclusion every year.

- Implications for equating
  - Building out over time is can considerably change the weight of the available content in terms of which elements of the dimensions are measured. Maintaining similarity is an important consideration in building out.
  - Similar issues to ELA where item sets are key. How to maintain reasonable representativeness of equating anchor items while constrained in large part to item sets?

An Inherent Tension: Focus vs. Transfer

- Balancing focus and transfer introduces considerable complexity:
  - Option 1: Curriculum and instruction don't address the *expectations*
    - They're the province of assessment
    - To address transfer, curriculum and instruction is based on novel *expectations* to embed transfer throughout the assessment
    - But there is that troubling aphorism that "what gets tested gets taught"
  - Additional equating challenges introduced by option 1
    - Since the domain of assessment is strictly defined by the existing expectations, there are no additional equating challenges introduced

An Inherent Tension: Focus vs. Transfer

- Option 2: Assessment does not address the *expectations*
  - They're the province of curriculum and instruction
  - To address transfer, the assessment is based on novel *expectations* addressing other cells in the domain matrix.
  - Enough novel *expectations* need to be developed with associated test content to avoid transferring the risk of narrowing the curriculum to a small set of *expectations* developed for assessment.
- Additional equating challenges introduced by option 2
  - Developing a set of novel expectations that is reasonably representative of the domain without touching the cells included in existing expectations.
  - Developing enough novel expectations that vary enough over forms to convey that the entire domain is eligible for assessment.

An Inherent Tension: Focus vs. Transfer

- Option 3: Both assessment and curriculum and instruction are primarily based on existing performance expectations
  - Core and matrix design
    - Core: existing expectations
    - Matrix: novel expectations slowly built out over years

- Additional equating challenges introduced by option 2
  - Since core is the same every year, for equating purposes, does not introduce any new implications

Thank You!