

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

Joseph A. Martineau
Senior Associate



Implication of Domain Structure: Sparse Domain Coverage

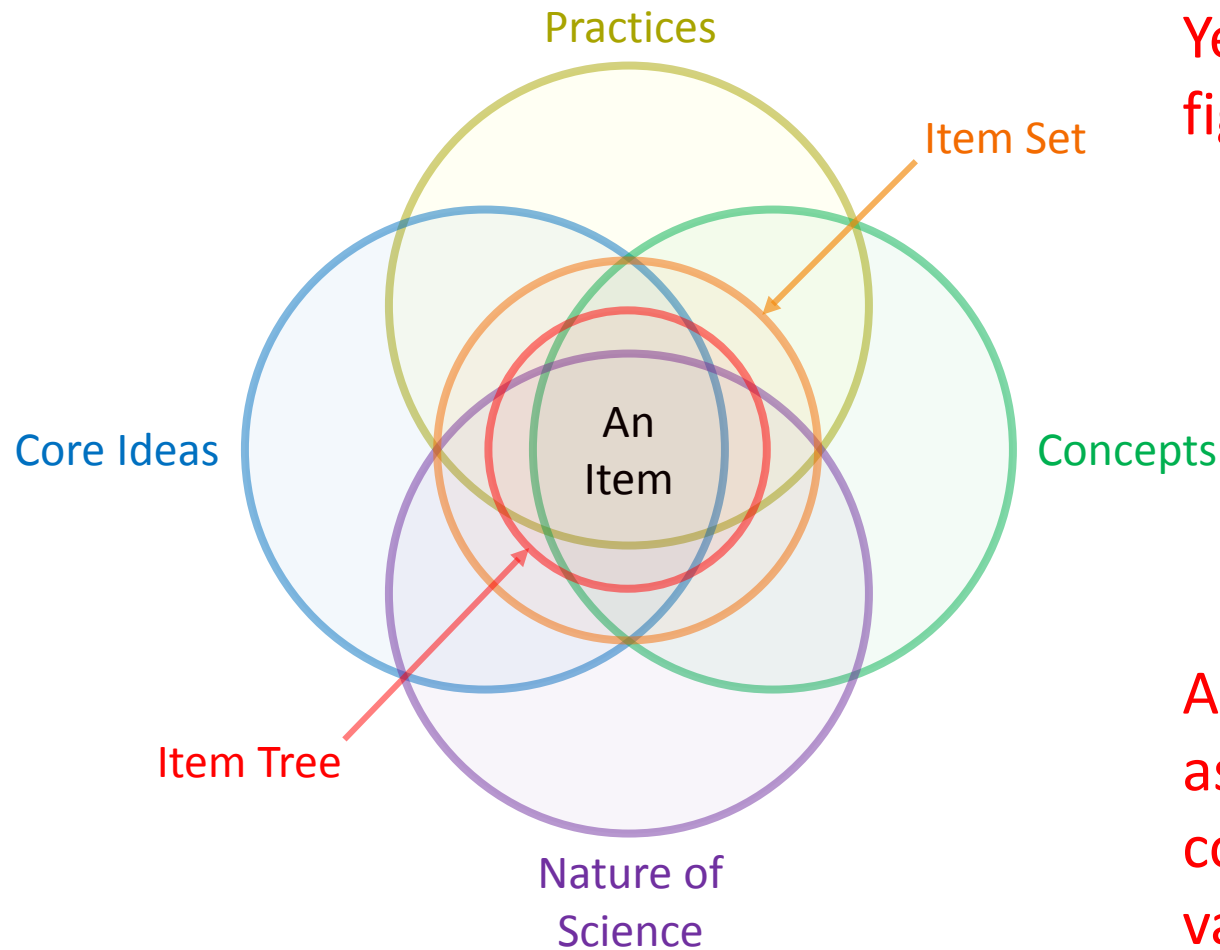


Dimension or Component	Type	Code	Element Coverage		
			3-5	MS	HS
<i>Disciplinary Core Ideas</i>	Dimension	DCI	91%	89%	93%
<i>Science & Engineering Practices</i>	Dimension	SEP	100%	100%	100%
<i>Crosscutting Concepts</i>	Dimension	CCC	71%	100%	100%
<i>Nature of Science</i>	Component	NOS	75%	75%	88%
Combination 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*



- 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



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!

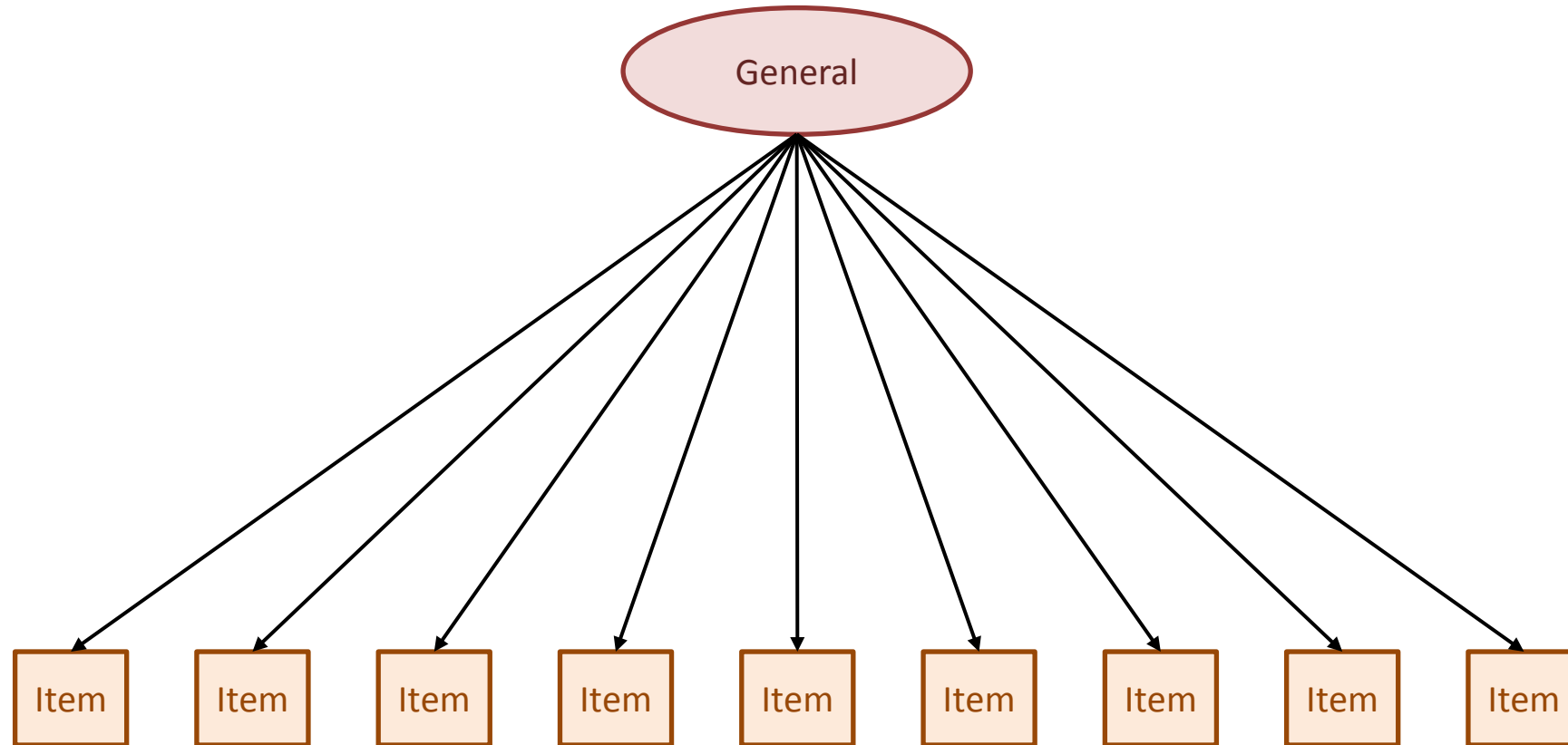
We need to stop trying to find subscores. We should be constructing them from the beginning.

–Nathan Dadey (late Wednesday night)



- 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 on a unidimensionally-modeled construct.
 - We can also report fewer subscores on a multidimensionally-modeled construct.

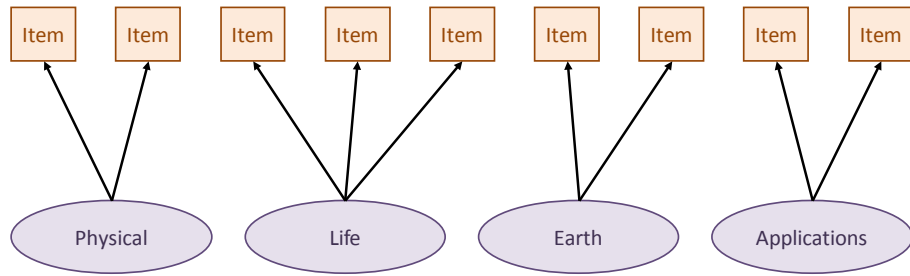
Dimensionality – Maybe Just This (Yay!)



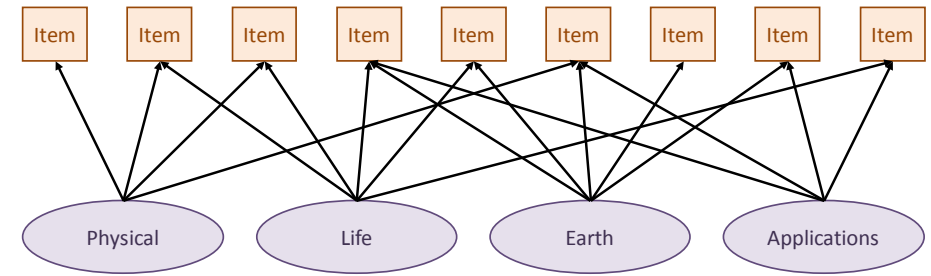
Dimensionality – Potential Scenarios



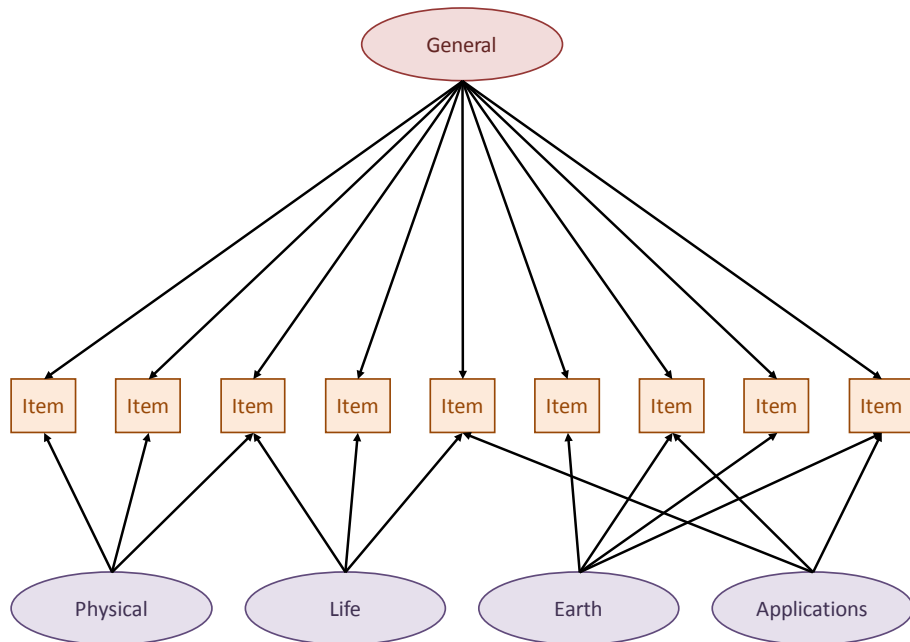
Simple Structure



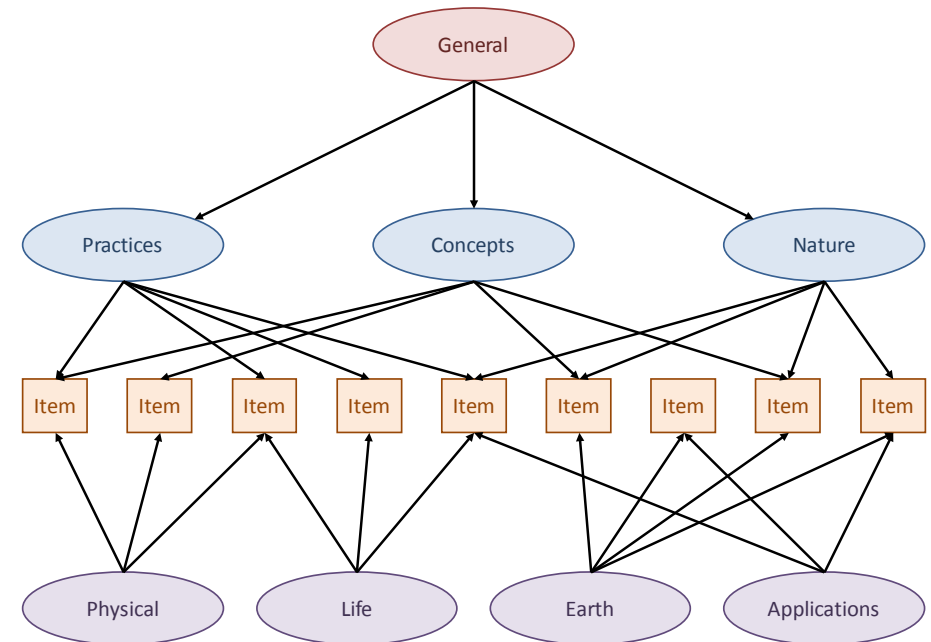
Complex Structure



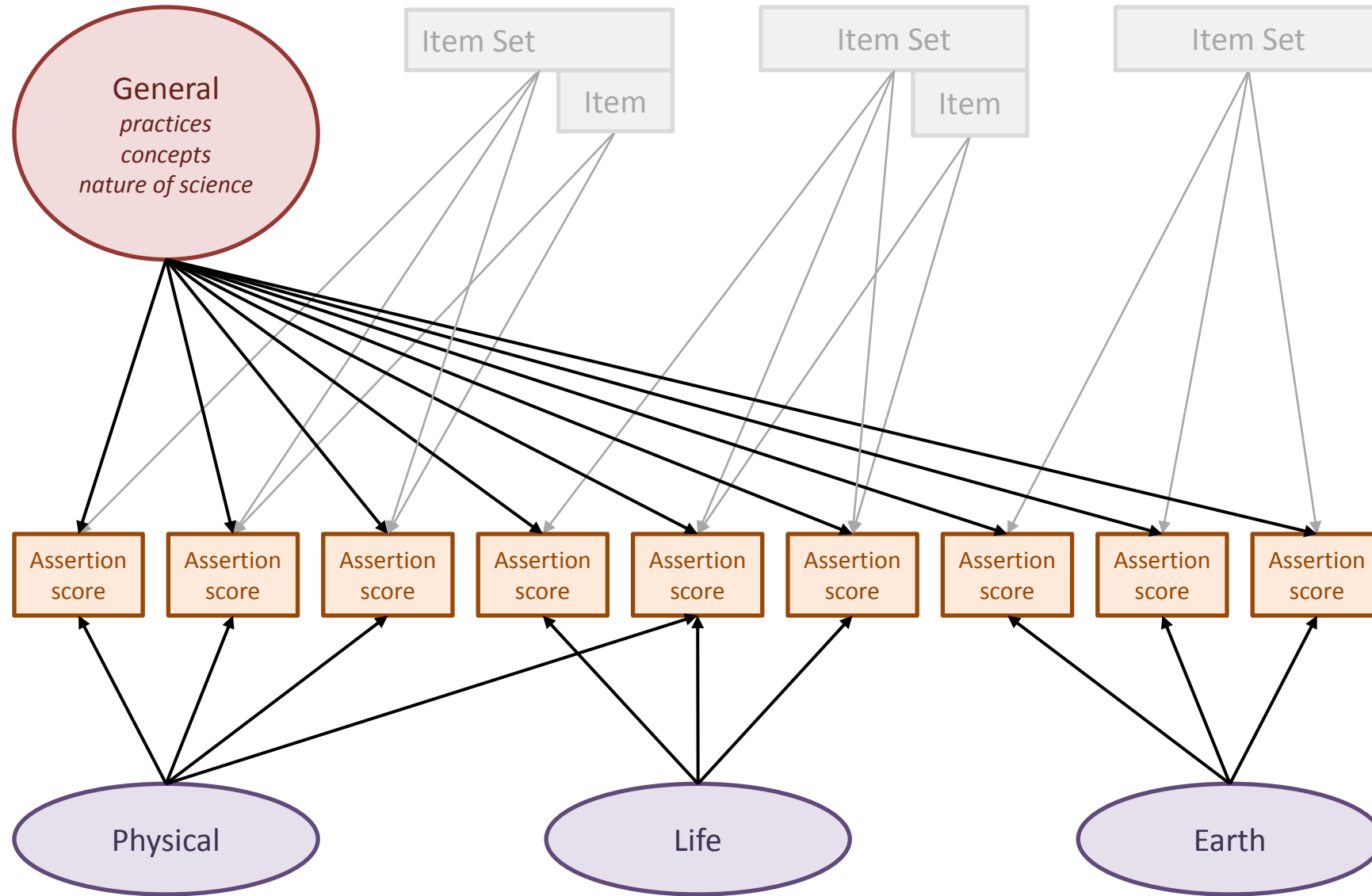
Bifactor



Higher Order



Dimensionality – My Somewhat Educated Guess





- 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.



- 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



- Representative equating item anchor sets
 - Similar issues to ELA where item sets are key
 - How to maintain reasonable representativeness while much of the available selection of items will be from item sets?
- 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
 - 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?



- 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



- 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.



- 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!

