

Claims, Evidence and Achievement Level Descriptors as a Foundation for Item Design and Test Specifications

Amy Hendrickson

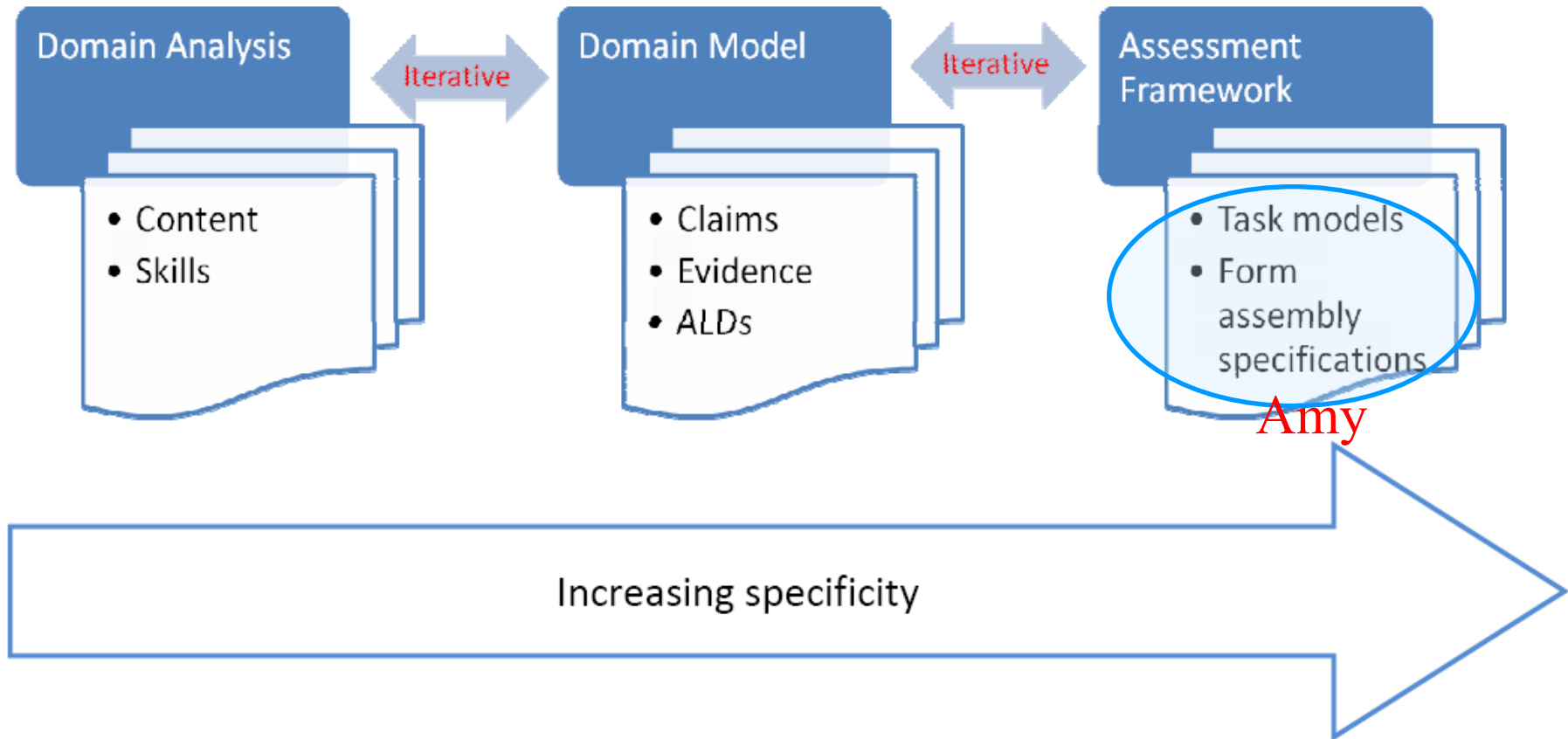
Kristen Huff

The College Board

Ric Luecht

University of North Carolina – Greensboro

This presentation



Purposes

- Discuss how claims, evidence and ALDs are used as input in the construction of the assessment framework.
 - Development of the task models
 - Development of test specifications
- Improved comparability and better supported score interpretations
- Examples from different disciplines
- Challenges and benefits of using ECD to construct the assessment framework for a large-scale assessment program

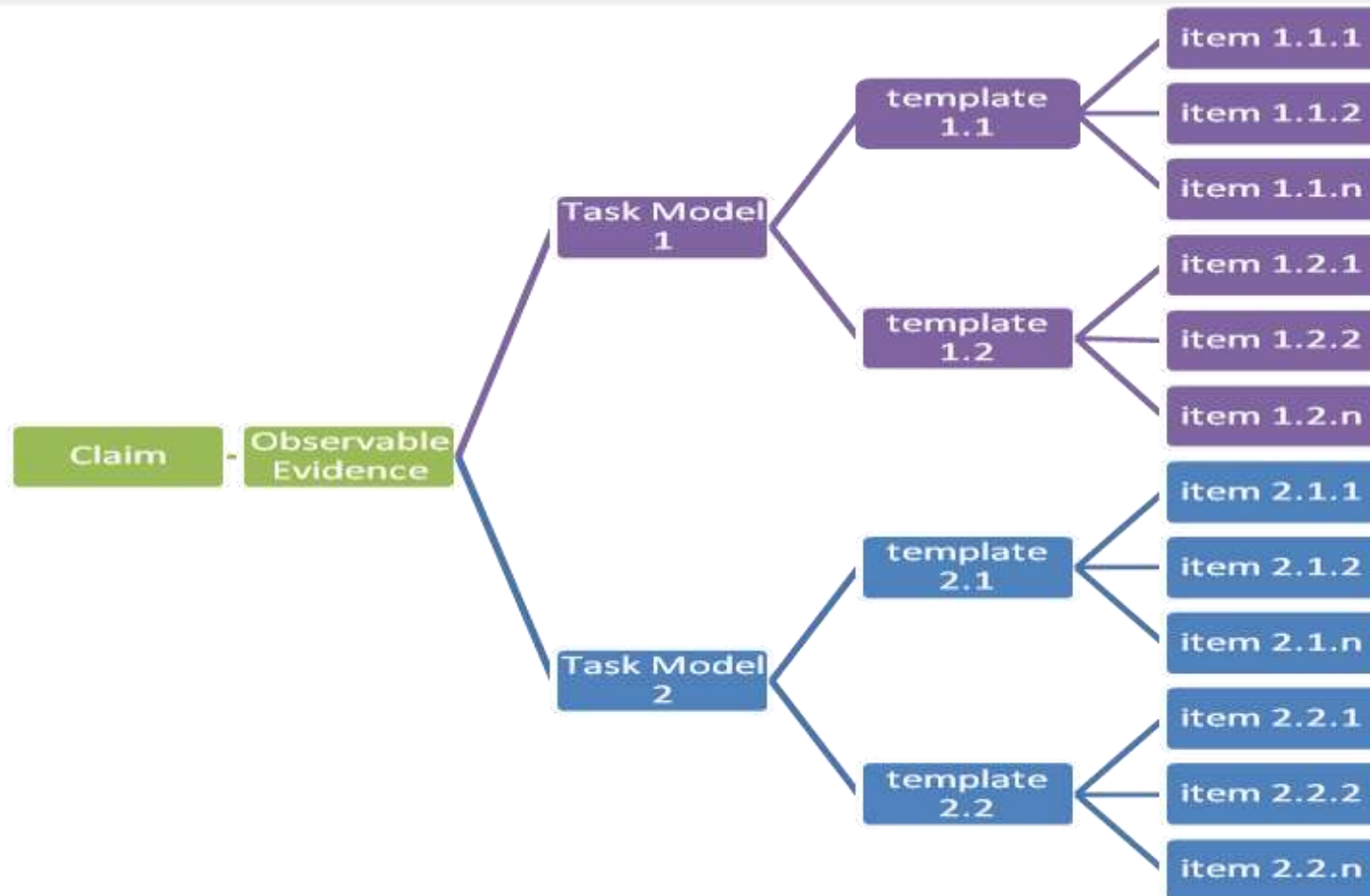
Task Models - The Basis for Item Design

- Conventional, non-ECD approaches
 - List of content and skills
 - Item format
 - Reviewed for adherence to the requirements and for fairness, edited as necessary
- ECD approaches
 - Design and development of task models
 - Provide the explicit link between the claims and evidence and the items
 - Support validity of score inferences

Task Models – Definition and Development

- Collection of relevant task features or variables
- Associated with a particular claim and evidence pair
- Multiple items, all providing essentially interchangeable evidence of achieving the claim
- Provide explicit guidance to item writers
- Process is iterative
- Flexibility and arbitrariness in number and degree of specificity

Sample Task Model Structure



Task Models – Considerations

- Decisions made jointly by assessment designers and item writers
- Prototype items
 - Inform specific features
 - Student response data helps inform decisions about features, variations and levels of specificity.
- Iteration between task models, templates, and items, and balance of expert judgment with student response data is important

Test Specifications – Conventional vs ECD

- Conventional approaches
 - Development of somewhat independent sets of test specifications
 - statistical specifications
 - content and skill specifications
 - May lead to scores with reasonable psychometric quality, but no support for the valid interpretation of student performance
- ECD approaches
 - An integrated set of specifications that include a clear articulation of claims to be made from test performance
 - Principled, replicable methods of gathering evidence to measure the ordered claims
 - Results in a psychometric scale that is consistent with the underlying construct/performance continuum

Test Specifications - Considerations

- Multiple inputs
 - Domain Model
 - Experts' ratings of importance of content and skills
 - Psychometric criteria
- Structure of the domain
 - Claims: skills-based versus integration of skills and content
 - Content relationships
 - Skill relationships
 - Content and skill relationships

Test Specifications - Development

1. Identify key variables
2. Determine the desired distributions of these variables
3. Merge the desired distributions
4. Ensure that the intended claims at each achievement level could be supported
 - a. Review distributions with domain experts
 - b. Modify distributions and domain model
5. Collect data
 - a. Make further refinements to the specifications

Test Specifications - Example from History

- European History, World History, and US History
- Variables
 - Historical thinking skills (interrelated and hierarchical)
 - Content
 - Themes (e.g., *Development and Interaction of Cultures*)
 - Periods (e.g., *Global Interactions, c. 1450 to c. 1750*)
 - Key concepts (e.g., *State Consolidation and Imperial Expansion*)
 - Geographical regions (e.g., *Europe*) – for World only
- Claims for the histories were skill based, even though content also plays an important role in the history exams
- Domain experts had to determine the weighting of the variables.

History Skill and Achievement Level Specifications Example

Skills	Skill Weights	Number of Items	ALD 3	ALD 4	ALD 5
Crafting Historical Arguments From Historical Evidence	25.00%	15	4-6	4-6	4-6
Historical argumentation	12.50%	7-8	1-3	1-3	1-3
Appropriate use of relevant historical evidence	12.50%	7-8	1-3	1-3	1-3
Chronological Reasoning	25.00%	15	4-6	4-6	4-6
Historical Causation	8.33%	5	1-2	1-2	1-2
Patterns of Continuity and Change Over Time	8.33%	5	1-2	1-2	1-2
Periodization	8.33%	5	1-2	1-2	1-2
Comparison and Contextualization	25.00%	15	4-6	4-6	4-6
Comparison	12.50%	7-8	1-3	1-3	1-3
Contextualization	12.50%	7-8	1-3	1-3	1-3
Historical Interpretation and Synthesis	25.00%	15	2-4	5-6	5-6
Interpretation	12.50%	7-8	1-3	1-3	1-3
Synthesis	12.50%	7-8	0-1	2-4	2-4
	100.00%	60	20	20	20

Task Models and Test Specifications - Challenges

- Domain model is first, but not only input
 - Sufficient time should be allotted for gathering domain expert ratings
 - Task models and test specifications may lead to domain changes
- Item coding
 - Generated and captured by the task models and used in the test specifications
 - Inter-related nature of the content features, skills, and achievement levels
 - items need to be coded for multiple instances of each variable
 - items allowed to satisfy one or more test specifications
 - Resource-intensive and requires sufficient infrastructure
- Must familiarize item writers with concepts of ECD

Task Models and Test Specifications - Benefits

- Items are generated from task models, which are derived directly from claims and evidence and are ordered according to achievement level
- Test specifications reflect the integration of content and skills required to distinguish student performance at various achievement levels
- The assessment framework integrates all of the artifacts from evidence-centered assessment design – the claims, evidence, and ALDs
- Thus, the assessment framework provides an operational synthesis of the evidentiary and validity argument for our claims about examinee proficiency

- Access this presentation online at <http://professionals.collegeboard.com/data-reports-research/cb/presentations>
- Please forward any questions, comments, and suggestions to:

Amy Hendrickson at:

ahendrickson@collegeboard.org