Principled Assessment Frameworks

Engineering the Future of Test Development

Matthew J. Burke, Ph.D.

May 15th, 2015

The future of testing is:

- Reliably predicting and controlling the difficulty of test items…

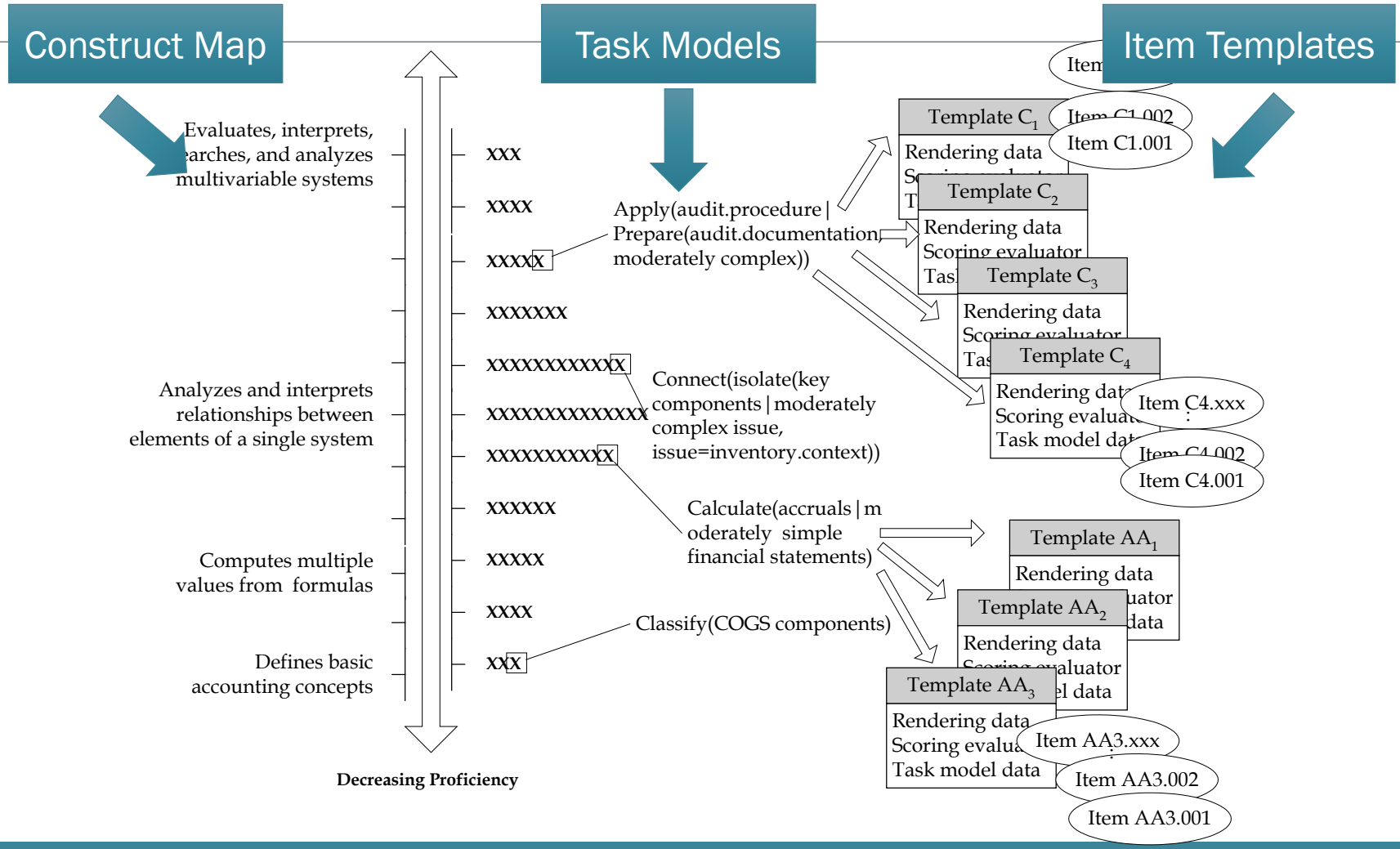
Assessment Engineering

- One of a class of principled assessment frameworks
  - Evidence-centered Design (Mislevy), Principled Design for Efficacy (Nichols), Principled Assessment Designs for Inquiry (IERI)
- Comprehensive, model-based view of test development, administration, and scoring
- Offers potential of both theoretical and practical improvements
  - Construct validity, Response processing validity
  - Item development, calibration, and scoring

Components of Assessment Engineering

- Construct Map
  - Visual representation of the score scale
  - Demarcates ordered proficiency claims relative to the scale
- Task Models
  - Aligned with the ordered proficiency claims
  - Each model represents a family of items providing comparable information
- Templates
  - Item rendering blueprints
  - Provide instructions for producing item isomorphs

Components of Assessment Engineering: Accounting Specific Example



Defining a taxonomy of skills

- Criteria of a cognitive taxonomy
  - Grain size, relevance, measurable, *hierarchical**
  - Revised Bloom's Taxonomy (Anderson et al., 2001)
- Distilling the requisite skills
  - Cognitive task analysis (CTA)
  - Reverse-engineering
  - Structure of the skills
    - *Hierarchical**, distinct, identifiable

➢Putting it all together
  - Incorporation into test specifications, guidance of practice analysis

AE: Modified Skill/Content Specification

Prepare financial documentation for reporting and presentation purposes in accordance with Reporting Framework (US GAAP/ IFRS).	
1. Balance sheet (UAS)	-*Understand disclosure requirements. (U)* -Identify information that needs to be disclosed. (U) -*Identify defining characteristics of accounting terms. (U)* - Categorize/ classify cash flow transactions, assets, liabilities, equity (U) -*Prepare adjusted trial balance. (A)* -Prepare supporting schedules or worksheets using accounting rules and procedures (A) -*Prepare financial statements (A)* - Calculate/ transcribe main components of financial statement (i.e. net income, total assets-liabilities-equity, cash flow categories) (A)
2. Income statement (UAS)	
3. Statement of comprehensive income (UAS)	
4. Statement of changes in equity (UAS)	
5. Statement of cash flows (UAS)	
6. Notes to financial statements (U)	
7. Consolidated and combined financial statements (UAS)	-*Prepare financial statements by combining information derived from a variety of sources (S)* -Prepare journal entries, worksheets or financial statement for consolidation with intercompany transactions (S)
8. First-time adoption of IFRS (U)	

Related Research

- Item difficulty modeling
  - Diehl, 2004; Embretson, 1998; Embretson and Daniel, 2008; Embretson and Gorin, 2001; Embretson and Wetzel, 1987; Gorin and Embretson, 2006
- Building/incorporating the infrastructure of AE
  - Luecht, 2015*; Luecht, 2013; Luecht, Burke and DeVore, 2009; Burke, DeVore, and Stopek, 2013; Burke and Stopek, 2013; Stopek and Burke, 2013; Burke, Stopek, and Eve, 2014; Furter, Burke, Morgan, and Kaliski, 2015
- Automatic item generation
  - Gierl, Lai, and Turner, 2012; Gierl and Lai, 2012; Alves, Gierl, and Lai, (2010); Gierl and Lai, ATP 2015*
- Automated test assembly
  - Van der Linden, 2006; Luecht, 1998
- Item family calibrations
  - Sinharay, Johnson, and Williamson, 2003; Glas and van der Linden, 2003; Geerlings, Glas, and van der Linden, 2011

Pros	Cons
-Confirmatory, model-based approach to test development -Strengthens validity argument -Directed item development -Decreased cost of test development in the long term -Reduced pre-testing demands -Standard setting/equating	-Extensive planning and preparation -Potential overkill in some assessment settings -Increased cost of test development in the short term -Requires niche experts in test development and modeling -Requires flexibility in pilot testing

Challenges

- Changing existing processes that work
- People are sometimes territorial
- Measurement concerns often follow practical and policy concerns
- Research is ongoing, work in progress
- No off the shelf products exist, must be custom made
- Doesn't work in every case*

➢ Establishing buy-in
  - ➢ Internal and external stakeholders
  - ➢ We are saying this will be better, but they need to come to that conclusion on their own.