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

- Evaluates, interprets, searches, and analyzes multivariable systems
- Analyzes and interprets relationships between elements of a single system
- Computes multiple values from formulas
- Defines basic accounting concepts

Decreasing Proficiency

**Task Models**

- Template C1
  - Rendering data
  - Scoring evaluator
  - Task model data

- Template C2
  - Rendering data
  - Scoring evaluator
  - Task model data

- Template C3
  - Rendering data
  - Scoring evaluator
  - Task model data

- Template C4
  - Rendering data
  - Scoring evaluator
  - Task model data

**Item Templates**

- Item C1.xxx
  - Item C1.002
  - Item C1.001

- Item C4.xxx
  - Item C4.002
  - Item C4.001

- Item AA3.xxx
  - Item AA3.002
  - Item AA3.001
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)
2. Income statement (UAS)  - Identify information that needs to be disclosed (U)
3. Statement of comprehensive income (UAS)  - Identify defining characteristics of accounting terms (U)
4. Statement of changes in equity (UAS)  - Categorize/classify cash flow transactions, assets, liabilities, equity (U)
5. Statement of cash flows (UAS)  - Prepare adjusted trial balance (A)
6. Notes to financial statements (U)  - Prepare supporting schedules or worksheets using accounting rules and procedures (A)
7. Consolidated and combined financial statements  - Prepare financial statements (A)
8. First-time adoption of IFRS (U)  - 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)
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

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