Evaluating New Approaches To Assessing Learning

Richard J. Shavelson, Min Li, Maria Araceli Ruiz-Primo & Carlos Cuauhtémoc Ayala

Keynote Address*
Joint Northumbria/EARLI Assessment Conference
University of Northumbria at Newcastle, Longhirst Campus
28 August 2002

*Available at http://www.stanford.edu/department/SUSE/SEAL
Overview

• Evaluating the Quality of Learning Assessments: Conceptual Framework
• Applying the Framework
• Concluding Comments
Conceptualizing Assessment: The Assessment Triangle

- **Cognition**: A model of that explains how students represent knowledge and develop competence
- **Observation**: Tasks or situations that prompt student to say, do, or create something to demonstrate knowledge
- **Interpretation**: A process for making sense of evidence

[Source: Pellegrino, Chudowsky & Glaser, 1999]
Evaluating Assessments: The Assessment Square Corners

- **Construct**: A working definition of what is to be measured
- **Assessment**: Systematic procedure for eliciting, capturing and scoring behavior
- **Observation**: Collecting and summarizing behavior in response to a task
- **Interpretation**: Inference from behavior on an assessment to the construct

[Sources: Ayala, Yin, Shavelson & Vanides, 2002; Ruiz-Primo, Shavelson, Li & Schultz, 2001]
Evaluating Assessments: The Assessment Square Analyses

The Analyses

- **Conceptual Analysis**: Identify domain of tasks and responses from construct definition.
- **Logical Analysis**: Logical evidence that task will evoke in a student a problem space and response consistent with construct.
- **Cognitive Analysis**: Empirical evidence on cognitive activities evoked by task/response.
- **Statistical and/or Qualitative Analysis**: Bring quantitative and/or qualitative data to bear on proposed assessment interpretation.

Construct Interpretation

Assessment Observation

Warranted Inference?
Applying Framework To TIMSS’ Achievement Test: Construct and Assessment

- The Construct: Science Achievement
- The Assessment: TIMSS Population 2 Science Test Items
  - Multiple-choice
  - Short-answer
- A combination of logical, cognitive and statistical analyses

[Source: Li (2001), Li & Shavelson, 2001]
Applying The Framework: Conceptual Analysis

- **Declarative**—knowing that
- **Procedural**—knowing how
- **Schematic**—knowing why
- **Strategic**—knowing when, where, and how to apply knowledge
Applying The Framework: Logical Analysis of TIMSS Items

- Task Demands
- Cognitive Demands—Assume competent 14 year old
- Item Openness
- Complexity
Logical Analysis: Item Coding System

• **Task Demands**: What does the item ask student to do?
  – Terms, symbols, vocabulary, definition
  – Procedures, steps, actions, algorithms
  – Models, relationships, explanation, principles

• **Cognitive Demands**: What prior knowledge and cognitive processes examinee may use and reason with?
  – Visualize
  – Calculate
  – Perform experiment
  – Recall information
  – Reason and interpret with models and principles
  – Plan and monitor behavior
  – Guess or eliminate wrong options
Logical Analysis:
Item Coding System (Cont’d.)

- **Item Openness**: How free in shaping item response?
  - Hands-on v. paper-and-pencil
  - Selected v. constructed response
  - Constrained v. open response
  - One v. multiple solution paths
  - Follow steps in instruction

- **Complexity**: How familiar, relevant, reading difficult is item?
  - Textbook vs. ill-structured task
  - Inclusion of irrelevant background information
  - Long, reading demanding descriptions and complicated vocabulary
  - Answers contradict everyday experience
Logical Analysis:
Declarative Knowledge Item

P6. What digestive substance is found in the mouth? What does it do?

- Assume *competent* 14 year old
- Task: Response expected to be a term, vocabulary (e.g., saliva), factual statement
- Cognitive Activity: Likely to be recall (question similar to form in which student learned content) with minimal reasoning to organize answer
- Openness: An open-ended question
Q11. Which statement explains why daylight and darkness occur on Earth?

A. The Earth rotates on its axis
B. The Sun rotates on its axis
C. The Earth’s axis is tilted.
D. The Earth revolves around the sun

- Task Demands: Asks for explanation “why”; a model can be used to answer
- Cognitive Demands: Requires reasoning with a model (unless memorized/recalled)
- Openness: The information forms a complete question that allows examinees to finish the item without reading alternatives
- Complexity: Reasonable reading load
Logical Analysis: Procedural Knowledge Item

P1. The graph shows the progress made by an ant moving along a straight line.

If the ant keeps moving at the same speed, how far will it have traveled at the end of 30 seconds?

A. 5cm  
B. 6cm  
C. 20cm  
D. 30cm

- Task Demands: Interpret diagram or apply algorithm
- Cognitive Demands: Apply the formula of Speed \( \div \) Distance or extend line
- Openness: Constrained—can work backwards from alternative
- Complexity: Moderate reading
Applying The Framework: Cognitive Analysis

• Assumptions:
  – Information processing model
  – Verbalization of working memory
  – Cognitive activity interpretation warranted

• Steps
  – Collect concurrent verbalization
  – Segment protocols
  – Code protocols
  – Analyze data
# Cognitive Analysis:  
Sequence Of Think-Aloud Study

<table>
<thead>
<tr>
<th>Session 1</th>
<th>Session 2-Session 4</th>
<th>Session 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction of the study, thinking aloud on the exercise problems-Shoe tying</td>
<td>Solving a group of multiple-choice and free-response items with thinking-aloud &amp; observations</td>
<td>Solving the two performance assessment tasks with thinking-aloud &amp; observations</td>
</tr>
<tr>
<td>5-10</td>
<td>10-15</td>
<td>15-30</td>
</tr>
<tr>
<td></td>
<td>2-5</td>
<td>5-10</td>
</tr>
</tbody>
</table>

[Source: Li, 2001]
Cognitive Analysis: Protocol Analysis

- Collect concurrent verbalizations—participants were instructed to verbalize *anything* while responding to test items
- *Segment* each participant’s verbal protocol—Li used the entire response to each item or task no matter how many statements or types of knowledge
- Code participants’ segments—Li developed a system that captured evidence the four types of knowledge.
- Examine coding consistency (reliability)
- Bring coded protocols data to bear on how participants employed different types of knowledge to represent and solve problems
Cognitive Analysis: Link Between Logical And Cognitive Analysis

Based on the knowledge-type construct of science achievement, we expected participants’ use of knowledge inferred from the protocols (cognitive analysis) to be congruent with the knowledge-types demanded by test items (logical analysis).

<table>
<thead>
<tr>
<th>Type of knowledge used</th>
<th>Pre-classified knowledge-type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Declarative</td>
</tr>
<tr>
<td></td>
<td>(n=9)*</td>
</tr>
<tr>
<td>Declarative</td>
<td>48</td>
</tr>
<tr>
<td>Procedural</td>
<td>0</td>
</tr>
<tr>
<td>Schematic</td>
<td>9</td>
</tr>
<tr>
<td>Strategic</td>
<td>2</td>
</tr>
</tbody>
</table>

*Number of responses: 48 = 9 x 6 participants
Chi-square = 208.12, p<.001
Based on the knowledge-type construct of science achievement, we expected the emerged factors from the item scores (statistical analysis) to be congruent with the knowledge-types in logical analysis.
Applying The Framework To Concept Maps: Construct and Assessment

• The Construct:
  – Knowledge structure (declarative)

• The Assessment
  – Concept-Map
  – A graph with nodes (concept terms) connected by labeled lines explaining relationship

[Source: Ruiz-Primo, Shavelson, Li, & Schultz, 2001]
The Study

• **Purpose:**
  – Compared “construct-a-map” and “fill-in-a-map” concept-mapping techniques
    • Fill-in-the-nodes with structure given
    • Fill-in-the-line labels with structure given
  – Motivation—fill in technique quicker, easier, and less expensive to score
  – Do both techniques provide equally valid measures of knowledge structure?

• **Domain:** Chemistry—atoms and molecules

• **Participants:**
  – Main study: 152 high-school students
  – Cognitive analysis *expert-novice* study:
    • 3 low-proficient students
    • 3 high-proficient students
    • 2 chemistry teachers

[Source: Ruiz-Primo, Shavelson, Li, & Schultz, 2001]
Applying The Framework: Conceptual Analysis

Concept-map Assessment of Knowledge Structure:

• Assumes long-term memory operates as a network with linked nodes

• Builds on empirical findings that experts’ knowledge, compared to novices’, is more:
  – Highly structured
  – Scientifically justifiable
Framework For Examining “Cognitive Validity” Portion Of Assessment Square

1. Cognitive activities evoked by assessment
2. Relation between cognitive activities and performance scores
3. Impact of variation in technique on cognitive activities
4. Correlation between scores on similar and different assessments

[Source: Ruiz-Primo, Shavelson, Li, & Schultz, 2001]
Logical Analysis: Concept-Map Directedness

Student Provides:
- Concept terms
- Linking Lines
- Linking Explanations
- Map Structure

Assessor Provides:
- Concept terms
- Linking Lines
- Linking Explanations
- Map Structure

[Source: Ruiz-Primo & Shavelson, 1996]
Cognitive Analysis: Think-Aloud

- **Micro-level Analysis** (Average reliability = 0.86)
  - Segment protocol into small units—phrases
  - Code phrases
    - Explanations ($N_2O_2$ is a molecular compounds because they are both nonmetals)
    - Monitoring (“I can’t remember exactly what this is“)
    - Conceptual errors (“Molecules are atoms”)
    - Inapplicable events (Reads instructions)

- **Macro-level Analysis** (Average: agreement 95%/Kappa .92)
  - Segment protocol by entire map
  - Code *planning* from beginning of protocol
  - Code *strategies* for entire protocol attending to sequence
Comparison Of Micro-Level Cognitive Activities

<table>
<thead>
<tr>
<th>Technique</th>
<th>n</th>
<th>Explanation</th>
<th>Monitoring</th>
<th>Conceptual Errors</th>
<th>No-Code Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct-a-map</td>
<td>8</td>
<td>39.08</td>
<td>28.22</td>
<td>9.78</td>
<td>22.91</td>
</tr>
<tr>
<td>Fill-in-the-nodes</td>
<td>8</td>
<td>4.64</td>
<td>40.93</td>
<td>0.16</td>
<td>54.27</td>
</tr>
<tr>
<td>Fill-in-the-lines</td>
<td>8</td>
<td>2.84</td>
<td>35.22</td>
<td>0.14</td>
<td>61.80</td>
</tr>
</tbody>
</table>

[Source: Ruiz-Primo, Shavelson, Li, & Schultz, 2001]
Comparison of Macro-Level Cognition: Strategy Sequence

- Select most general concept (1.1)
- Select concept related to general concept (2.1)
- Monitor performance by reading propositions (3.1)
- And so on…

[Source: Ruiz-Primo, Shavelson, Li, & Schultz, 2001]
Link Between Cognitive Activities and Performance Scores

Correlations Between Verbal Proportion Scores and Performance Scores by Category (n=7) *

<table>
<thead>
<tr>
<th>Mapping Technique</th>
<th>Category Explanation</th>
<th>Category Monitoring</th>
<th>Category Conceptual Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct-a-map</td>
<td>.33</td>
<td>-.29</td>
<td>-.27</td>
</tr>
<tr>
<td>Fill-in-nodes</td>
<td>.12</td>
<td>-.31</td>
<td>a</td>
</tr>
<tr>
<td>Fill-in-lines</td>
<td>.01</td>
<td>-.83^b</td>
<td>a</td>
</tr>
</tbody>
</table>

* Outlier deleted on each correlation
a No variability in one of the variables.
b Correlation is significant at the .05 level.

[Source: Ruiz-Primo, Shavelson, Li, & Schultz, 2001]
Comparison Of Scores On Mapping Techniques

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Cognitive Analysis</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct-A-Map</td>
<td>Cognitive Processes</td>
<td>Performance</td>
</tr>
<tr>
<td>Fill-In-A-Map</td>
<td>Cognitive Processes</td>
<td>Performance</td>
</tr>
</tbody>
</table>

Mean Scores and Standard Deviations Across the Mapping Techniques

<table>
<thead>
<tr>
<th>Mapping Technique</th>
<th>Max</th>
<th>Students</th>
<th>Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n = 6</td>
<td></td>
</tr>
<tr>
<td>Construct-a-map</td>
<td>1</td>
<td>0.63</td>
<td>0.31</td>
</tr>
<tr>
<td>Fill-in-the-nodes</td>
<td>12</td>
<td>10.67</td>
<td>1.97</td>
</tr>
<tr>
<td>Fill-in-the-lines</td>
<td>12</td>
<td>8.83</td>
<td>3.13</td>
</tr>
</tbody>
</table>

[Source: Ruiz-Primo, Shavelson, Li, & Schultz, 2001]
Comparison Of Scores On Mapping Techniques

Mean Scores and Standard Deviations Across the Mapping Techniques

<table>
<thead>
<tr>
<th>Mapping Technique</th>
<th>CM</th>
<th>NOD (n = 5)</th>
<th>LIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct-a-map</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fill-in-the-nodes</td>
<td>.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fill-in-the-lines</td>
<td>.91</td>
<td>.42</td>
<td></td>
</tr>
</tbody>
</table>

[Source: Ruiz-Primo, Shavelson, Li, & Schultz, 2001]
Concluding Comments

*The assessment square:*

- Provides a framework for examining new learning assessments
- Integrates conceptual, logical cognitive and statistical analyses of assessment-interpretation claims