On Formative (Embedded) Science Assessment

Richard J. Shavelson
Stanford University

Great Minds Program Dialogue

Science Education in the 21st Century: Pushing the Envelope on Student Assessment
April 11, 2001
Overview

• National policy trends and assessment
• New and continuing developments in assessment of science achievement
• New focus on classroom assessment
• Concluding comments
National Policy Trends in Science Education

• Continued commitment to inquiry-based science-education reform for all students from AAAS, NAS and NSF
• Increased pressure for improvement in science teachers’ pedagogical content knowledge (Glenn)
• Increased pressure for “hard” evidence of the positive impact and to “scale up” innovations that work
• Increased accountability with traditional tests
Increased Accountability With Traditional Tests

**Type**
- Formative
- Summative

**Purpose**
- Learning
- Certification
- Accountability

**Agency**
- Student
- Teacher
- Individual
- External tests
- Sample surveys

National Focus

Paul Black 3/98
Developments in Assessment: Conceptual Framework

- **Declarative Knowledge** (Knowing the “that”)
  - Domain-specific content: facts, concepts, principles

- **Procedural Knowledge** (Knowing the “how”)
  - Domain-specific production systems

- **Schematic Knowledge** (Knowing the “why”)
  - Problem schemata, mental models

- **Strategic Knowledge** (Knowing the “which” and “when”)
  - Domain-specific strategies, heuristics, contexts

**Proficiency**
- Extent (How much?)
- Structure (How is it organized?)
- Others (Precision? Efficiency? Automaticity?)

**Cognitive Tools:**
- Planning
- Monitoring
Developments in Assessment: Framework Links to Tests

<table>
<thead>
<tr>
<th>Extent</th>
<th>Declarative Knowledge</th>
<th>Procedural Knowledge</th>
<th>Schematic Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td>Multiple-Choice</td>
<td>Performance Assessments</td>
<td>Performance Assessments</td>
</tr>
<tr>
<td></td>
<td>Fill-in</td>
<td>Journals</td>
<td>Interviews</td>
</tr>
<tr>
<td></td>
<td>Journals</td>
<td></td>
<td>M-C Tests</td>
</tr>
<tr>
<td>Others</td>
<td>Concept Maps</td>
<td>Procedure Maps</td>
<td>Models/ Mental Maps</td>
</tr>
<tr>
<td></td>
<td>Semantic Maps</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Developments in Assessment: Example Multiple-Choice

Air is made up of many gases. Which gas is found in the greatest amount?

A. Nitrogen
B. Oxygen
C. Carbon Dioxide
D. Hydrogen

TIMSS Population 2
Developments in Assessment: Example Concept Map

From White & Gunstone: “Probing Understanding” (1992, p. 16)
Developments in Assessment: Example Performance Assessment

Students are asked to model the path of the sun from sunrise to sunset and use direction, length, and angles of shadows to solve location problems.
Developments in Assessment: Example Mental Model Problem

• (A) A rocket is moving along sideways in deep space, with its engine off, from point A to point B. It is not near any planets or other outside forces. Its engine is fired at point B and left on for 2 sec while the rocket travels from point B to point C. Draw in the shape of the path from B to C. (Show your best guess for this problem even if you are unsure of the answer.)

• (B) Show the path from C after the engine is turned off on the same drawing.

### Developments in Assessment: Validity Evidence from TIMSS-R

<table>
<thead>
<tr>
<th>Knowledge Type</th>
<th>Percent Multiple Choice</th>
<th>Percent Open Ended</th>
<th>Percent Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic (Declarative &amp; Schematic)</td>
<td>10</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>Procedural</td>
<td>6</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Conceptual (Declarative)</td>
<td>16</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>Factual (Declarative)</td>
<td>38</td>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td>“Not Science”</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>73</strong></td>
<td><strong>27</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>
Developments in Assessment: Validity Evidence

Correlations from Shultz’s Dissertation
(N=109 6th Graders Studying Ecology)

- Reading and Multiple-Choice: 0.69
- Reading and Concept Map: 0.53
- M-C and CM: 0.60

- Reading and Performance Assessment: 0.25
- M-C and PA: 0.33
- CM and PA: 0.43
New Focus on Classroom Assessment

<table>
<thead>
<tr>
<th>Type</th>
<th>Purpose</th>
<th>Agency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formative</td>
<td>Learning</td>
<td>student, teacher</td>
</tr>
<tr>
<td>Summative</td>
<td>Certification</td>
<td>external tests, individual</td>
</tr>
<tr>
<td></td>
<td>Accountability</td>
<td>external tests, sample surveys</td>
</tr>
</tbody>
</table>

Paul Black 3/98

California focus

Match-Mismatch
The “Reverse” Logic of Developing Assessments for a Unit of Instruction
Assessment Task Structure

Content Knowledge

Process Skills

Rich

Constrained

Open

Lean

More Cognitive Demanding

Less Cognitive Demanding

Baxter & Glaser, 98
Classroom (Formative) Assessment

• *Everyday teaching practice conceived as integral in assessment*
• Assessment used to determine *gap* between what student knows and knowledge goal
• Teacher, *peer*, and *self* assessments comprise classroom assessment
• Feedback critical to close the gap
  – Grades?
  – Qualitative feedback useful to closing gap?
  – Both?
Classroom Assessment: Examples

• Teacher
  – Observations
  – Questions
  – Interviews
  – Journals
  – Curriculum-provided and/or teacher assessments

• Self and Peer (*Clear goals absolutely essential*)
  – Review/grade each other’s work
  – Review each other’s journals
  – Reflect on learning
Types of Embedded Assessments

- Multiple-Choice
- Short Answer
- Completion
- Exercises
- Venn diagrams

- POE
- Interview
- Portfolio
- Drawings
- Others
The Gap

Use formative assessment information to:

- Determine the gap between student performance and the learning goals
- Provide *qualitative* feedback to student on how to close the gap
Journals: An Assessment Tool for Teachers and Students

**Journals:**

- Are a written account of what students do in their science class, and possibly, of what they learn.
- May provide an *unobtrusive* indicator of class experiences.
- Are seen as an *immediate* assessment -- in very close proximity to the curriculum.
- Are viewed as assessments at two levels:
  - *at the individual level* are considered a source of evidence bearing on student’s performance over a course of instruction.
  - *at the classroom level* are a source of evidence of opportunities students had to learn science.
Some Findings About Students’ Journals as Assessment Tools

**Reliability:**
Raters can consistently identify journal entries
Students’ science journals can be reliably scored

**Validity**
Inferences about implementation using journal scores were justified
Inferences about students’ performance were also encouraging

**Usefulness**
Unit implementation and teacher feedback scores helped to explain differences in the performance across classrooms
Conclusions

• Classroom formative assessment may be a critical key to improving students’ achievement in inquiry science

• Assessment reform will require extensive use of technology to manage time and cost*

• Collaboration between researchers and teachers absolutely essential*
Technology & Our Collaboration: Cost Reduction and Scoring Efficiency

• Computer adaptive testing--testing extent of declarative and (perhaps) strategic knowledge
• Computer concept and cognitive mapping--testing structural knowledge
• Computer simulation of an investigation--testing procedural and/or strategic knowledge
Collaboration--Convergence of Hands-On and Computer Simulation PAs

![Diagram with correlation coefficients]

- $r_{H1H2} = .53$
- $r_{C1C2} = ?$
- $r_{H1C1} = .52$
- $r_{H2C2} = ?$
- $r_{H1C2} = ?$
- $r_{C1H2} = .45$
- $r_{C1C2} = ?$
Thanks for Inviting Me!