On Developing A Comprehensive Science Assessment Plan for Delaware

Richard J. Shavelson
Stanford University

May 14, 2001
The performance measure ("output") is an indicator of the desired behavior, not the behavior itself.

In business, the output measure (e.g., revenue or stock price) is a very close proxy to valued outcomes. It guides business decisions and actions. You can’t manage a business if you can’t measure it’s outcome.

In education, outcomes are many and debated. The outcome indicator, most often a multiple-choice achievement test, is but a proxy for the desired outcome. When this indicator becomes an end in itself, and it does in education, well-intentioned accountability may very well distort the system it was intended to improve.

Source: March (1994)
Education Indicator System
Framework for Cognitive Outputs

(Knowing the “that”) (Knowing the “how”) (Knowing the “which,” “when,” and “why”)

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

Domain-specific content:
- Facts
- Concepts
- Principles

Domain-specific production systems

Problem schemata/strategies/operation systems

Cognitive Tools:
Planning
Monitoring
Air is made up of many gases. Which gas is found in the greatest amount?

A. Nitrogen

B. Oxygen

C. Carbon Dioxide

D. Hydrogen
Assessment of Declarative Knowledge
Structure: Eleven-Year-Old’s Concept Map

From White & Gunstone: “Probing Understanding” (1992, p. 16)
Assessment of Procedural Knowledge: 
Performance of a Daytime Astronomy Investigation

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.
Procedural Knowledge: Performance Assessment: TIMSS Pop. 2

**PULSE**

**At this station you should have**
- A watch
- A step on the floor to climb on

**Read ALL directions carefully.**

**Your task:**
Find out how your pulse changes when you climb up and down on a step for 5 minutes.

**This is what you should do:**
- Find your pulse and be sure you know how to count it. IF YOU CANNOT FIND YOUR PULSE ASK A TEACHER FOR HELP
- Decide how often you will take measurements starting from when you are rest.
- Climb the step for about 5 minutes and measure your pulse at regular intervals.

1. Make a table and write down the times at which you measured your pulse and the measurements you made.
2. How did your pulse change during the exercise?
3. Why do you think your pulse changed in this way?
Assessment of Strategic Knowledge: Mental Models

• (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.

Strategic Knowledge: Mental Models

What causes day and night?

A. The earth spins on its axis (.66)
B. The earth moves around the sun (.26)
C. Clouds block out the sun (.03)
D. the sun goes round the earth (.04)

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>Total</td>
<td>73</td>
<td>27</td>
<td>100</td>
</tr>
</tbody>
</table>
# Linking Assessments to Achievement Components

<table>
<thead>
<tr>
<th>Extent</th>
<th>Declarative Knowledge</th>
<th>Procedural Knowledge</th>
<th>Strategic Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Multiple-Choice</td>
<td>• Performance Assessments</td>
<td>• Multiple-Choice</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>• Performance Assessment</td>
</tr>
<tr>
<td>Structure</td>
<td>• Concept Maps</td>
<td>Procedure Maps</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Cognitive Maps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CLAS: Aligning Instruction & Accountability

A. “On Demand” Matrix Sampling of Tasks & Tests
B. Standardized Curriculum-Embedded Assessments
C. Portfolios

Student’s Score
Conceptual Framework For CLAS

A. Matrix Sample Benchmark: Multiple-Choice & Performance-Based Assessment

"Moderated" Score: Individual, School & District Score

Teacher Moderation

Sample from Class for Aggregation

Teacher Calibration & Professional Development

B. Standardized Curriculum-Embedded Assessments

C. Portfolios

Sample from Class for Aggregation

Teacher Calibration & Professional Development

Individual Level of Performance

Aggregate Level of Performance
Timeline For Implementation

A. Standardized Benchmarks

B. Curriculum Embedded Assessments &

C. Portfolios

1990 _______________ Increasing Teacher Responsibility for Assessment _______________ 2000
Caveat: Governor Wilson’s Rationale For Discontinuing CLAS

SB 1273 [CLAS] takes a different approach . . . .
Instead of mandating individual student scores first, with performance-based assessment incorporated into such scores as this method is proven valid and reliable, it mandates performance-based assessment now and treats the production of individual student scores as if it were the experimental technology--which it clearly is not. In short, SB 1273 stands the priority for individual scores on its head.
British Task Group on Assessment and Testing (TGAT): Recommendations

Assessment

- Results should give direct information about pupil achievement in relation to objectives (criterion-referenced criterion)
- Results should provide diagnostic information for pupil learning needs (formative-evaluation criterion)
- Scales should be comparable across classes and schools (calibrated/moderated criterion)
- Scales should be set up to trace pupil’s developmental trajectories over grade levels (progression criterion)

[Sources: Black (1966), Daugherty (1995)]
Ken Baker warmly welcomed the report. Whether he had read it properly I do not know: if he had it says much for his stamina. Certainly I had no opportunity to do so before agreeing to its publication . . . that it was then welcomed by the Labour party, the National Union of Teachers and the Times Educational Supplement was enough to confirm for me that its approach was suspect.

Source: Black (1996, p. 5)