Why Assessment Matters?

Jonathan Osborne
School of Education
Stanford University

Feb 11, 2011
Main Points

1. What counts matters but not all that matters can be counted

2. The intentions of the curriculum are read from the assessment items not the curriculum documents

3. Assessments operationalize constructs

4. Knowing what’s wrong matters as much as knowing what is right.
What can not be measured?

- Persistence
- Curiosity
- A critical disposition
- A commitment to evidence as the basis of belief
- Awe and Wonder
- Imagination
Assessments Operationalize Constructs

- Intentions are read from the test items
- Items become the embodiment of what matters
- Bloom’s Taxonomy of Cognitive Skills
  - Recall
  - Comprehension
  - Application
  - Analysis
  - Synthesis
  - Evaluation
Cognitive Demand: National Tests at Age 14
Cognitive Demand: National Tests at Age 16

- **Recall**: High
- **Comprehension**: Moderate
- **Application**: High
- **Analysis**: Moderate
- **Synthesis**: Low
- **Evaluation**: Very low
Effect of High Stakes Assessment

- Narrowing of the Curriculum & Teaching to the Test
- Increasing Fragmentation
- More Teacher Centred Pedagogy

PISA

**Context**
Life situations that involve science and technology. *(See Figure 3.2)*

**Competencies**
- Identify scientific issues
- Explain phenomena scientifically
- Use scientific evidence *(See Figure 3.3)*

**Knowledge**
- About the natural world (knowledge of science)
- About science itself (knowledge about science) *(See Figures 3.4 & 3.5)*

**Attitudes**
How you respond to science issues:
- interest
- support for scientific enquiry
- responsibility

How you do so is influenced by:
### TIMSS Framework 2011

<table>
<thead>
<tr>
<th>Knowing (35%)</th>
<th>Applying (35%)</th>
<th>Reasoning (30%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Recall</td>
<td>• Compare, Contrast, Classify</td>
<td>• Analyze</td>
</tr>
<tr>
<td>• Define</td>
<td>• Use Models</td>
<td>• Integrate/Synthesize</td>
</tr>
<tr>
<td>• Describe</td>
<td>• Relate</td>
<td>• Hypothesize/Predict</td>
</tr>
<tr>
<td>• Illustrate with Examples</td>
<td>• Interpret Information</td>
<td>• Design</td>
</tr>
<tr>
<td>• Demonstrate Knowledge of Scientific Instruments</td>
<td>• Find Solutions</td>
<td>• Draw Conclusions</td>
</tr>
<tr>
<td></td>
<td>• Explain</td>
<td>• Generalize</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Evaluate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Justify</td>
</tr>
</tbody>
</table>
Three Forms of Knowledge in Science

Content

Procedural

Epistemic
Procedural Knowledge

Concepts of Evidence
- Error
- Measurement
- Instrumentation
- Reliability/Validity
- Variables/Fair Testing

Epistemic Knowledge

Ideas-About-Science
- The role of peer review
- Observation
- Theory
- Model
- Hypothesis
The Absence of Critique?

Construction v Critique
A Sample Question

Janet was asked to do an experiment to find how long it takes for some sugar to dissolve in water. What advice would you give Janet to tell her how many repeated measurements to take?

- A. Two or three measurements are always enough
- B. She should take 5 measurements
- C. If she is accurate she only needs to measure once
- D. She should go on taking measurements until she knows how much they vary
- E. She should go on taking measurements until she gets two or more the same
Bad Science
Ben Goldacre

'A fine lesson in how to skewer the enemies of reason and the peddlers of cant and half-truths.'
The Economist

'You'll laugh your head off, then throw all those expensive health foods in the bin.'
Observer Book of the Year

The Sunday Times top ten bestseller
Some Concluding Thoughts

1. Assessments need to be more flexible

2. The variance in assessments should be large?
Some Concluding Thoughts

1. Assessments need to be more flexible

2. The variance in assessments should be large?

3. In the trade off between reliability and validity, validity is what matters.