Assessing and Accounting for Student Achievements: The Quest to Hold Higher Education Accountable

Richard J. Shavelson
Stanford University

Invited Seminar
Carnegie Foundation for the Advancement of Teaching
October 3, 2000
Overview of Talk

• Motivation for and overview of study
• Accountability for achievements
• **Assessment of achievements**
  – Conceptual framework
  – Examples of cognitive assessments
  – Role of technology
  – Links with Carnegie
• Areas for collaboration
Motivation for Study

• Increased demand for accountability
  – New York’s Report Card
  – Virginia’s audit
  – Developments in England, New Zealand, Australia, Hong Kong

• Lessons from K-12 education: Benefits & Costs
  – Benefits include increased content achievement primarily in basic skills (Tennessee, Texas) and some teacher responsiveness
  – Costs include narrowing educational goals, reduced flexibility, teaching to the test, and cheating

• Problematic application to education--outputs are often distal proxies for desired outcomes
Overview of Study

• Historical, political, social and conceptual background for study
  – Precollege: beginning with common school movement
  – College: beginning with land grant institutions
• Framework for and case studies of assessing learning:
  – Cognitive
  – Civic responsibility and other “non-cognitive”
• Framework for and cases studies of accountability
• Options for alternative accountability systems from the decision maker’s perspective
Accountability and Assessment

- **Accountability** is a procedure by which a polity’s (citizen, politician, public manager, or client) acts to have public agencies account for the resources they use and the values they create.

- **Assessment** is “theory-driven” measurement and/or description of variables that provide data on or describe inputs (resources), processes (use of resources), outputs (products) and outcomes (valued consequences).

- An **accountability system** is a routine, systematic, “theory-driven” effort open to public debate intended to:
  - Collect data on 2 or more organizations
  - Transform those data into information relevant to evaluating performance
  - Transmit this information to some audience external to the organization through scores (often ratings or rankings) and sometimes (case) descriptions
Framework for Thinking About Accountability

• Locus of information bearing on accountability
  – Inputs (resources)
  – Processes (use of resources)
  – Outputs (direct products such as achievement scores)
  – Outcomes (valued consequences)

• Criteria for evaluating accountability systems
  – Validity (Fidelity of output assessment(s) to desired outcomes)
  – Comprehensivenessness (includes relevant variables)
  – Comprehensibility (to potential users)
  – Relevance (to needs of potential user)
  – Reasonableness (demands on organization)
  – Functionality (leads to appropriate behavior)

Source: Gromley & Weimer, 1999
Models of Output (Achievement) Accountability Systems

- **Absolute Standard:**
  Performance of a system is measured against some internal or external standard of minimally acceptable (or highly respectable) level of performance.

- **Relative Standard:**
  - **Value-Added** where a system’s performance is compared against its expected performance given the nature of its inputs.
  - **Time-Series** that monitors system indicators (e.g., graduation rates, achievement scores) over time.
  - **Internal Audit** that links assessment of learning with the teaching and learning mission of the institution, with an externally verifiable internal quality-control mechanism.
  - **External Audit** that ties a system’s funding to ranking of indicators such as graduation rates, retention rates, and faculty teaching and research productivity.

- **Approximation Standard:**
  Model that evaluates a system against predictors of a system’s outcomes over time such as active learning, student-faculty interaction, and student time on task.
Actors in the Accountability Drama

Top Down Accountability

Citizens

Politicians

Public Managers

Service Providers

Corporate Purchasers

Clients

Government Purchasers

Bottom Up Accountability

Source: Gromley & Weimer (1999)
Framework for Tracking Consequences

Accountability Information

Consumers
(Prices)

Manager
resources
prestige

Overseers
(Budgets)
(Discretion)

Organizational Flexibility

Functional
Process Improvement
Input Reallocation
Managerial Focusing
Mission Enhancement
***

Dysfunctional
Self Selection
Cream Skimming
Teach to Test
Deception
Blame Messenger

Source: Gromley & Weimer (1999)
Assessment:
Framework for Cognitive Outputs

Declarative Knowledge
(Knowing the “that”)

Procedural Knowledge
(Knowing the “how”)

Strategic Knowledge
(Knowing the “which,” “when,” and “why”)

Proficiency
Low High

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
Assessment of Declarative Knowledge: Multiple-Choice--TIMSS Pop. 2

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.
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.

## 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>Structure</td>
<td>• Multiple-Choice</td>
<td>Performance</td>
<td>• Performance</td>
</tr>
<tr>
<td></td>
<td>• Fill-in</td>
<td>Assessments</td>
<td>Assessments</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Interviews</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• M-C Tests</td>
</tr>
<tr>
<td>Others</td>
<td>Concept Maps</td>
<td>Procedure Maps</td>
<td>Models/</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mental Maps</td>
</tr>
</tbody>
</table>
Some Empirical Evidence on Links between Knowledge and Measurement Methods

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
Role of Technology: 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
Links with Carnegie: The Carnegie Teaching Academy

• What are the assessment practices of faculty in the sciences, social sciences and humanities? (Linda Suskie of AAHE says no good studies! She and Ted Marchese studied practices considered exemplary by North Central--very few!)

• What are the assessment practices of the Carnegie fellows (infer from their studies of teaching)? Source of case studies?

• What do you all know about assessment practices?
Areas for Collaboration

• Assessment practices of Carnegie Fellows
• Assessment practices more generally
• Case studies
• Role of assessment of learning and teaching in accountability models
• *Ongoing advice greatly appreciated*!!!