Assessing Student Learning: The Quest To Hold Higher Education Accountable

Rich Shavelson

CASBS Seminar
December 13, 2000
Overview Of Talk

• Motivation for studying higher education accountability
• Sketch of envisioned study
• Institutional and state accountability systems ("report cards")
• Criteria for evaluating report cards
• Questions for us to address
Motivation for Study

• Respond to increased demand for accountability as conceived by, for example:
  – New York’s Report Card
  – Virginia’s audit
  – Developments in England, New Zealand, Australia, Hong Kong

• Avoid K-12’s negative consequences experiences: Benefits & costs
  – Benefits include increased content achievement (primarily in basic skills—e.g. Tennessee, Texas) and teacher responsiveness
  – Costs include narrowing educational goals, reduced flexibility, teaching to the test, and cheating

• Develop design principles to reduce transfer of inappropriate conceptions of accountability to higher education—outputs are often distal proxies for desired outcomes
The Question: How Can The Public Get Control Over Higher Education?

Rising costs, part-time faculty, non-traditional students, and for-profit institutions have fueled concern about higher education:

“The days when most public officials and their constituents viewed higher education as a innate good deserving of public moneys, with or without measurable outcomes, are over…. And today it is not possible in the public sector in South Carolina and in many other states to spend taxpayers’ hard-earned money without accounting for how it is spent, sometimes in detail…”

-- Rayburn Barton, Commissioner of Higher Education, South Carolina
The Response: Accountability

“One of the prime tools of effective private sector management is an accountability system that includes clear goals, a well-designed incentive structure and solid performance measures. Building this kind of system into American education is a fine idea. But we have to recognize that the development of accurate education measurements represents an enormous challenge”

-- Jim Thompson, President of The RAND Corporation, justifying the timing of policy pieces on the Texas Assessment System in the LA Times
Academia’s Reaction: Problems

- SUNY stops and studies proposed system-wide student comprehensive achievement test
- System-wide Faculty Senate asserts that:
  - Faculty responsible for general education design, implementation, & assessment
  - Campus differentiation fundamental
  - Assessment should be campus centered
  - Assessment design requires faculty representation
- System-wide Faculty Senate resolves that:
  - Provost should suspend his system-wide uniform comprehensive test of student achievement in the first two years of college
  - Faculty and others should develop a plan for a campus-based student outcomes assessment program
Nutshell Of New Study

• Historical, political, social and conceptual background for study
  – Precollege: beginning with common school movement
  – College: beginning with land grant institutions and post WWII education benefits

• Framework for and case studies of accountability

• Framework for and case studies of assessing learning:
  – Cognitive
  – Civic responsibility (etc.)

• Options for alternative accountability systems from the decision maker’s perspective
Accountability And Assessment

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

• 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 organizations through scores (often ratings or rankings) and sometimes (case) descriptions

• **Assessment** is “theory-driven” measurement (and description) of indicators that characterize inputs (resources), processes (use of resources), outputs (products) and outcomes (valued consequences)
The Complex Accountability Stage

Accountability Demand

Top Down
- Citizens
- Politicians
- Bureaucrats

Information
- Students
- Parents
- Government & Corporate Purchasers

Bottom Up

Accountability Context

International
- Econ. Competition
- H.E. Demand
- Acct. Cases

National
- Student Aid
- Research
- Accreditation

Region/State
- Accreditation
- Accountability

Non-Government Accountability Suppliers
- US News
- Zemsky (Mkt. Seg.)
- PPHE

Higher Education Institution Mission
Inputs --> Process--> Outputs--> Outcomes
- Humanities & Sciences
- General Education & Departments/Programs

Inputs
- Historical
- Social
- Economic
- Political
- Judicial

Outputs
Accountability System Models

- **Absolute Standard:**
  - Performance of a system is measured against some internal or external standard of minimally acceptable (or highly respectable) level of performance (e.g., NAEP)
  - Internal Audit that links assessment of learning with the teaching and learning mission of the institution, with an externally verifiable internal quality-control mechanism (e.g., Colorado)

- **Relative Standard:**
  - Value-Added where a system’s performance in producing learning is compared against its expected performance given the nature of its inputs (Tennessee)
  - Time-Series that monitors system indicators over time (e.g., graduation rates, achievement scores)
  - External Audit that ties a system’s funding to ranking of indicators such as graduation rates, retention rates, and faculty teaching and research productivity (South Carolina)

- **Approximation Standard:**
  A model that evaluates a system against known predictors of a system’s outcomes over time such as active learning, student-faculty interaction, and student time on task (NSSE)
Ranking Colleges: South Carolina’s Performance Funding*

Inputs (Resources)

- Class Size & Student/Faculty Ratio (Ave. 30-35 in universities; 16-21 in tech colleges)

Processes (Resource Use)

- Average hours taught by full-time teaching faculty (Financial incentives to increase)

Outputs (Direct Products)

- Graduation rate (2 year degrees in 3 and 4 year degrees in 6)

Outcomes (Goals)

- Student Employment (Alum surveys & state employment data)

Faculty:

- Credentials (Accred. agency criteria)
- Performance Review (Lose $ if not follow Southern accrediting agency standards)
- Post-tenure review (as recommended by accrediting agency)
- Compensation (Salaries => nat’l ave.)
- Availability to Students (Anon. eval by students)

- Percent of full-time employees who are faculty members (29.6 for 4-year colleges & 40.1 for 2-year)
- Accreditation of degree programs by recognized bodies
- Inst. Emphasis on Teacher Ed Quality & Reform (4-year: accredit., student performance on nat’l tests, % minority grads in academic disciplines w. teacher shortages)

- Percent who pass certification exam (differs by sector)
- Percent grads who continue education (Enroll within 3 years)
- Credit hours earned by grads (Avoid more credits than needed for graduation)

*In process
Ranking Colleges: Institutional Report Cards

Inputs (Resources)
- SAT/ACT Scores
- HS GPA/Class Rank
- Selectivity
- Financial Resources
- Percent Students Out-of-Province
- Per Pupil Expenditure

Processes (Resource Use)
- Class Size
- Student/Faculty Ratio
- Frosh Retention
- Faculty Reputation
- % Full-time Faculty
- Percent Enroll Part Time
- Ratio BA/BS to Total Undergraduate Enroll
- Percent 1st-Year Classes Taught by Tenured Faculty
- $CN spent on Student Services
- Libraries

Outputs (Direct Products)
- Graduation rate
- Alumni Giving

Outcomes (Goals)

Non-Governmental Accountability Suppliers:
- US News & World Report
- Zemsky’s Market Mapping
- Both US News & Zemsky
- MacCleans
- Both US News and MacCleans
- All Three
New Contenders: Institutional & State Report Cards

Inputs (Resources)
- Preparation
- Participation
- Affordability

Processes (Resource Use)
- Persistence & Completion
- Student Report of:
  1. Academic challenge
  2. Active & collab learning
  3. Interaction with faculty
  4. Enrichment (ed abroad)
  5. Support for social life

Outputs (Direct Products)
- Educational Gains & Returns to State
- Student learning
- Student reported gains toward personal goals and satisfaction

Outcomes (Goals)

- National Center for Public Policy and Higher Education’s State Report Card
- National Survey of Student Engagement
Criteria For Evaluating Accountability Systems

- **Validity** (Fidelity of output assessment(s) to desired outcomes)
- **Comprehensiveness** (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)
Validity Issues: Especially With Learning Outputs

- Accountability must be inferred from observing outputs in any system where all actions cannot be observed directly.
- To do this “inferencing,” the performance measure 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)
## Higher Education Outcomes

<table>
<thead>
<tr>
<th>Goal</th>
<th>% “Absolutely Essential”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense of maturity &amp; manage on own</td>
<td>71</td>
</tr>
<tr>
<td>Ability get along with people different from self</td>
<td>68</td>
</tr>
<tr>
<td>Improved problem-solving &amp; thinking ability</td>
<td>63</td>
</tr>
<tr>
<td>Learning high-tech skills</td>
<td>61</td>
</tr>
<tr>
<td>Specific expertise &amp; knowledge in chosen career</td>
<td>60</td>
</tr>
<tr>
<td>Top-notch writing &amp; speaking</td>
<td>57</td>
</tr>
<tr>
<td>Responsibilities of citizenship</td>
<td>44</td>
</tr>
<tr>
<td>Exposure to great writers and thinkers</td>
<td>32</td>
</tr>
</tbody>
</table>

Source: J. Immerwahr for National Center for Public Policy and Higher Education (October 2000)
# Alverno College Criterion-Performance Approach

## Academic Year/Level

<table>
<thead>
<tr>
<th>“Abilities”</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>Self-assesses</td>
<td>Analytic</td>
<td>Uses Com Techniq</td>
<td>Integr Comm Abil</td>
</tr>
<tr>
<td>Analysis</td>
<td>Observe</td>
<td>Infers</td>
<td>Relates</td>
<td>Integrates</td>
</tr>
<tr>
<td>Problem Solving</td>
<td>Self-assesses</td>
<td>Defines Prob</td>
<td>Resolves Prob</td>
<td>Implements &amp; Eval Sol</td>
</tr>
<tr>
<td>Valuing in Decision Making</td>
<td>ID Values</td>
<td>Infers Impl. Val.</td>
<td>Relates val to technol</td>
<td>Appl Val. Proc</td>
</tr>
<tr>
<td>Social Interaction</td>
<td>Self-assesses</td>
<td>Analyses Grps</td>
<td>Eval Self &amp; Grp</td>
<td>Perf Effect in Grps</td>
</tr>
<tr>
<td>Global Perspective</td>
<td>Self-assesses</td>
<td>Exam cplx relatls</td>
<td>Exam Mult Persp</td>
<td>Resp to Loc/Glob Iss</td>
</tr>
<tr>
<td>Effective Citizenship</td>
<td>Self-assesses</td>
<td>Dev Strats Inform Resp</td>
<td>ID Org Str Ach Goal</td>
<td>Design Strat</td>
</tr>
<tr>
<td>Aesthetic Responsiveness</td>
<td>Artic Psnl Resp</td>
<td>Explain Psnl Resp</td>
<td>Relates work to Ctxt</td>
<td>Makes/Defend Qual Judg</td>
</tr>
</tbody>
</table>
Truman College: Value-Added Approach

<table>
<thead>
<tr>
<th>Who Takes It?</th>
<th>Nationally Normed</th>
<th>Surveys</th>
<th>Qualatative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman</td>
<td>• College Assessment of Academic Proficiency or • Academic Profile</td>
<td>Cooperative Institutional Research Program (UCLA) Freshman Week Survey</td>
<td>• Academic Profile • Freshman Interview • Project</td>
</tr>
<tr>
<td>Sophomore</td>
<td></td>
<td>Institutional Student Survey</td>
<td>Sophomore Writing Experience</td>
</tr>
<tr>
<td>Junior</td>
<td>• CAAP or • AP</td>
<td>ISS</td>
<td>• AP • Junior Interview • Project</td>
</tr>
<tr>
<td>Senior</td>
<td>Senior Test in Major (GRE)</td>
<td>Graduating Student Questionnaire</td>
<td>• Capstone Course • Portfolio</td>
</tr>
<tr>
<td>Alumni</td>
<td></td>
<td>• Alumni Survey • Employer Surve</td>
<td></td>
</tr>
<tr>
<td>Faculty</td>
<td></td>
<td>Faculty Survey</td>
<td></td>
</tr>
</tbody>
</table>
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
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: Assessments and Knowledge 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, • Fill-in</td>
<td>Performance Assessments</td>
<td>• Performance Assessments, • Interviews, • M-C Tests</td>
</tr>
<tr>
<td>Others</td>
<td>Concept Maps</td>
<td>Procedure Maps</td>
<td>Models/Mental Maps</td>
</tr>
</tbody>
</table>

- **Declarative Knowledge**
- **Procedural Knowledge**
- **Strategic Knowledge**
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
Many Questions... But For A Start

• What social science theories provide useful lenses for thinking about accountability?
• What kinds of incentives need to be built in?
• What exemplary assessment-of-achievement practices might be incorporated?
• What should be avoided to reduce negative consequences?