Using Data for Enhancement: Don’t Let Proving Trump Improving

Alexander C. McCormick, Ph.D.
Center for Postsecondary Research
Indiana University Bloomington, USA

Surveys for Enhancement
London
4 June 2015
Agenda

- Disclosure
- Setting the context
- Unpacking accountability and enhancement
- Measurement matters
- Some examples
- Implications for enhancement
- Discussion
First, some linguistic clarifications...

“Assessment”
“Improvement”
“Faculty”
Context

Mounting pressure for accountability and transparency

- Rankings and league tables
- Cost escalation
- Value concerns
The New Public Management

- From “accountability for process” to “accountability for results”
- Proliferation of formal programs that seek to ratchet up attention to results in higher education
  - “Results” defined how, by and for whom?
Context, continued

New focus on institutional assessment

- Movement toward cultures of evidence
  - Evidence-informed decision making
- Demands for accountability and consumer information
- Engaging the quality discourse
  - Counterpoint to rankings (NSSE, UKES)
  - Imports from the management world (TQM, CQI, Learning organizations)
  - Scholarship of teaching & learning
Accountability

What have you done for me lately?

- Exchange relations
- Funders want proof of value for money
  - Government agents
  - Students & parents
  - Alumni
  - Governing Boards
Accountability, continued

In case you haven’t noticed...

HIGH STAKES
The Design Challenge

- The key design challenge of a performance-based quality regime is not selecting the right measures (a technical problem)
- The challenge is ensuring that the system motivates the right behavior (an alignment problem)
- This is more difficult than it seems!
The Proxy Problem

- Social scientists frequently use proxy measures to “stand in” for an abstract or diffuse construct
- Performance measurement systems and rankings rely heavily on proxy measures
- This substitution of part for whole is assumed to be unproblematic
The Proxy Problem (continued)

Quality domain

Subdomain of interest

Measurable with error

Measurable without error

These zones not measured

Adapted from Bevan & Hood, 2006
The Proxy Problem (continued)

Overall educational quality

Formal learning

Tested knowledge

Grades, graduation

Adapted from Bevan & Hood, 2006

These zones not measured
Behavioral Consequences

- When high stakes are attached to a proxy measure, decision makers focus on improving the measure rather than the broader attribute it represents.
- "Gaming" responses are more likely in high-stakes scenarios.

A system of performance rewards linked to precise measures is not an incentive to perform well; it is an incentive to obtain a good score.

March, 1984
UK National Health Service

**Targets for emergency room wait times**

- Large discrepancies between reported times and patient survey responses
- Delaying patient admission to manipulate when the clock starts
- Increasing staffing, canceling surgeries during the performance monitoring period
UK National Health Service

**Targets for ambulance response time**
- Record of “corrected” times showing a large spike at the target time

**Targets for outpatient waiting time**
- Cancellation and delay of appointments that do not figure in the metric
UK National Health Service

*Targets for patients to see a general practitioner within 48 hours*

- Refusals to schedule appointments more than 48 hours in advance

Examples from Bevan & Hood, 2006
US K-12 Education

“No Child Left Behind” law that sanctions schools based on student test scores

- Manipulating attendance on test days
- Several high-profile cases of systematic fraud
US Higher Education

**US News rankings**

- “Reputation management” to improve external ratings
- Increased emphasis on entrance scores in admissions
- Test score reporting anomalies
- Increased reliance on early decision (guaranteed to enroll) to improve admissions yield
US Higher Education

**Accountability systems that emphasise retention & graduation rates**

*The following are anecdotal*

- Reducing the number of risky admits
- Pressuring faculty not to fail students
Carnegie Classification

(The following are anecdotal)

- Targeting funding increases to fields that are represented in the criteria for research funding
Some choice quotes...
"Goodhart’s Law"  
(Goodhart, 1975)

Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes.

Simplified as...  
(Elton, 2004)

When a measure becomes a target, it ceases to be a good measure.
"Campbell’s Law"

(Campbell, 1979)

The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.
Implications:

Proving vs. Improving
Proving vs. Improving

For the entity being held accountable, the process can lead to an emphasis on PROVING rather than IMPROVING.
Proving vs. Improving

**Proving** suggests...

- Strategy
- Advocacy
- Damage control
- Image management
- Influencing the process

**Improving** suggests...

- Reflection
- Analysis
- Candid examination
- Diagnosis
- Intervention / action
## Two Paradigms of Assessment
(Ewell, 2009)

<table>
<thead>
<tr>
<th>Strategic Dimensions</th>
<th>Improvement</th>
<th>Accountability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intent</strong></td>
<td>Formative (Improvement)</td>
<td>Summative (Judgment)</td>
</tr>
<tr>
<td><strong>Stance</strong></td>
<td>Internal</td>
<td>External</td>
</tr>
<tr>
<td><strong>Predominant Ethos</strong></td>
<td>Engagement</td>
<td>Compliance</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Application Choices</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instrumentation</strong></td>
<td>Multiple/Triangulation</td>
<td>Standardized</td>
</tr>
<tr>
<td><strong>Nature of Evidence</strong></td>
<td>Quantitative and Qualitative</td>
<td>Quantitative</td>
</tr>
<tr>
<td><strong>Reference Points</strong></td>
<td>Over Time, Comparative, Established Goal</td>
<td>Comparative or Fixed Standard</td>
</tr>
<tr>
<td><strong>Communication of Results</strong></td>
<td>Multiple Internal Channels and Media</td>
<td>Public Communication</td>
</tr>
<tr>
<td><strong>Uses of Results</strong></td>
<td>Multiple Feedback Loops</td>
<td>Reporting</td>
</tr>
</tbody>
</table>
## Two Paradigms of Assessment (Ewell, 2009)

### Table 1: Two Paradigms of Assessment

<table>
<thead>
<tr>
<th>Strategic Dimensions</th>
<th>Improvement</th>
<th>Accountability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intent</strong></td>
<td>Formative (Improvement)</td>
<td>Summative (Judgment)</td>
</tr>
<tr>
<td><strong>Stance</strong></td>
<td>Internal</td>
<td>External</td>
</tr>
<tr>
<td><strong>Predominant Ethos</strong></td>
<td>Engagement</td>
<td>Compliance</td>
</tr>
</tbody>
</table>

### Application Choices

<table>
<thead>
<tr>
<th></th>
<th>Assessment for Improvement Paradigm</th>
<th>Assessment for Accountability Paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instrumentation</strong></td>
<td>Multiple/Triangulation</td>
<td>Standardized</td>
</tr>
<tr>
<td><strong>Nature of Evidence</strong></td>
<td>Quantitative and Qualitative</td>
<td>Quantitative</td>
</tr>
<tr>
<td><strong>Reference Points</strong></td>
<td>Over Time, Comparative, Established Goal</td>
<td>Comparative or Fixed Standard</td>
</tr>
<tr>
<td><strong>Communication of Results</strong></td>
<td>Multiple Internal Channels and Media</td>
<td>Public Communication</td>
</tr>
<tr>
<td><strong>Uses of Results</strong></td>
<td>Multiple Feedback Loops</td>
<td>Reporting</td>
</tr>
</tbody>
</table>
Accountability requires the entity held accountable to demonstrate, with evidence, conformity with an established standard of process or outcome. The associated incentive for that entity is to look as good as possible, regardless of the underlying performance. Improvement, in turn, entails an opposite set of incentives. Deficiencies in performance must be faithfully detected and reported so they can be acted upon.

Peter Ewell (2009)
Thought Experiment #1

1. Your university uses a popular voluntary assessment tool that provides valid comparisons between institutions.

2. Results are disappointing but suggestive, and there’s interest on campus.

3. Relations with policy makers are frosty.

4. New accountability mandate: *Thou shalt publish thy assessment results.*

5. What happens?
Thought Experiment #2

1. Your university participates in a widely used student survey focused on teaching & learning that provides valid comparisons between institutions.

2. A national publication gains access to results and uses them in a ranking.

3. Students discover that the rankings are based on their survey responses.

4. What happens?
The Challenge for Enhancement

- Don’t let proving trump improving
  - Refocus accountability as professional responsibility
  - Emphasise formative purposes
  - Resist “chasing the measure”

- Stay alert and ever vigilant!
Discussion
References


