Multiple Measures in the Context of Performance-Based Accountability

Brian Stecher
Reidy Interactive Lecture Series
September 23, 2011
Performance-Based Accountability

• Link incentives with measured performance to improve the attainment of goals

• Key elements of PBA systems:
  – Consensus on goals
  – Measures
  – Target structure
  – Incentives

• Multiple measures might improve the performance of the system
No Child Left Behind

Incentives
- Identification
- Corrective Action
- Restructuring

Measures
- Performance level on mathematics test
- Performance level on reading test
- Graduation rate
- Attendance rate
- Test-taking rate

Target Structure
- Increasing targets to 100% (AMO) by subgroup
- Local targets
- Fixed target (95%)

Parent reporting
Highly qualified teacher
Twenty States Currently Report Measures Beyond NCLB

• Outcome indicators included in ratings
  – Additional tested subjects (e.g., history, social studies)
  – College readiness (e.g., ACT, AP)
  – Growth in performance
  – Weighted test scores along whole performance range

• Additional indicators
  – Safe and Supportive Environment
  – Students at-risk of failure
  – Interim assessments/advanced course-taking
Multiple Measures and PBA

- PBA Theory of Action: Measured performance related to endorsed goals triggers incentives that shape behavior thus promoting goals
- How might multiple measures contribute?
  - More accurate information improves the signal
  - More complete information aligns better with goals
  - More robust information reduces corruption
  - More sensitive targeting heightens responsiveness
Multiple Measures:
More Accurate Information
Multiple Measures: More Accurate Information (2)
Multiple Measures: More Complete Information
Multiple Measures: More Complete Information

Omni Parker House ★★★★★
Hotel photos | Map this hotel | Amenities

#25 of 73 hotels in Boston

1,186 reviews

"Good stay" Aug 16, 2011
"Great Location, Stay here!" Aug 16, 2011

Show Prices
Rooms available from under $150
More Complete Information: Society Has Diverse Goals for Education

- Academic Skills
- Practical and Fine Arts
- Health and Nutrition
- Social Skills
- Psychological Skills
- Citizenship
More Complete Information: Many Ways to Measure These Diverse Goals

- Academic skills
- Practical and Fine Arts
- Health and Nutrition
- Social Skills
- Psychological Skills
- Citizenship

- Tests
- Performances
- Access & Participation
- Stakeholder Surveys
- Direct Observation
- Administrative Records
- Post-Secondary Data
Multiple Measures: More Robust Information

- Campbell’s Law
  - 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.
  - In the context of high stakes testing this leads to score inflation

- Multiple indicators diffuse the pressure of accountability on indicators.
Multiple Measures: More Sensitive Targeting

Ratings E-book readers

In performance order, within types. (Types designated A, B, etc.)

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Rank</th>
<th>Brand &amp; model</th>
<th>Price</th>
<th>Overall score</th>
<th>Test results</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>REGULAR-SIZED MODELS</strong> All have 6-inch e-ink screens, measured diagonally, except as noted.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Barnes &amp; Noble Nook Simple Touch Reader</td>
<td>$140</td>
<td>78</td>
<td></td>
<td>60 days 7.5</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Amazon Kindle 3G + Wi-Fi</td>
<td>$190</td>
<td>77</td>
<td></td>
<td>60 days 8.2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Sony Reader Daily Edition PRS-950SC</td>
<td>$300</td>
<td>73</td>
<td></td>
<td>10-22 days 9.6</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>Amazon Kindle Wi-Fi</td>
<td>$140</td>
<td>73</td>
<td></td>
<td>60 days 7.7</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>Amazon Kindle Wi-Fi with Special Offers</td>
<td>$115</td>
<td>73</td>
<td></td>
<td>60 days 7.7</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>Barnes &amp; Noble Nook Color</td>
<td>$250</td>
<td>72</td>
<td></td>
<td>Up to 8 hr. 15.8</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>EnTourage Pocket eDiGe</td>
<td>$320</td>
<td>66</td>
<td></td>
<td>7 hr. 23.3</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>Sony Reader Touch Edition PRS650SC</td>
<td>$230</td>
<td>64</td>
<td></td>
<td>2 weeks 7.4</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>Barnes &amp; Noble Nook 1st Edition 3G + Wi-Fi</td>
<td>$170</td>
<td>63</td>
<td></td>
<td>10 days 12.0</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>Barnes &amp; Noble Nook 1st Edition Wi-Fi</td>
<td>$120</td>
<td>60</td>
<td></td>
<td>10 days 11.4</td>
</tr>
</tbody>
</table>
Multiple Measures: More Sensitive Targeting

School Content Area Weights for the Most Common Grade Spans, K–8

<table>
<thead>
<tr>
<th>Content Area</th>
<th>2010–11 API Test Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K–5</td>
</tr>
<tr>
<td>CST/CMA/CAPA in ELA</td>
<td>56.5%</td>
</tr>
<tr>
<td>CST/CMA/CAPA in Mathematics</td>
<td>37.6%</td>
</tr>
<tr>
<td>CST/CMA/CAPA in Science, Grades 5 and 8</td>
<td>5.9%</td>
</tr>
<tr>
<td>CST in History–Social Science, Grade 8</td>
<td>--</td>
</tr>
</tbody>
</table>

School Content Area Weights for Grades 9–12

<table>
<thead>
<tr>
<th>Content Area</th>
<th>2010–11 API Test Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9–12</td>
</tr>
<tr>
<td>CST/CMA/CAPA in ELA, Grades 9–11</td>
<td>27.1%</td>
</tr>
<tr>
<td>CST/CMA/CAPA in Mathematics, Grades 9–11</td>
<td>18.1%</td>
</tr>
<tr>
<td>CST/CMA/CAPA in Science, Grades 9–11</td>
<td>22.9%</td>
</tr>
<tr>
<td>CST in History–Social Science, Grades 9–11</td>
<td>13.9%</td>
</tr>
<tr>
<td>CAHSEE ELA, Grades 10–12</td>
<td>9.0%</td>
</tr>
<tr>
<td>CAHSEE Mathematics, Grades 10–12</td>
<td>9.0%</td>
</tr>
</tbody>
</table>
Challenge: Measuring Performance that Matters

Long-Term Outcome

- College Success
- Career Success
- Healthy Living
- Citizenship
Challenge: Measuring Performance that Matters

Intermediate Outcome

Long-Term Outcome

- College Readiness
- Career Readiness
- Course-Taking
- Athletic Participation
- Student Government
Challenge: Measuring Performance that Matters

- AP Course Offering
- College Counseling
- Art/Music Courses
- Model UN
Challenge: Measuring Performance that Matters

- Qualified Teachers
- Art/Music Facilities
- Community Partners
Challenge: Balancing Benefits and Costs/Risks

- Breadth ↔ Focus
- Complexity ↔ Transparency
- Comprehensiveness ↔ Affordability
- Uniformity ↔ Flexibility
- Formative ↔ Summative Purposes
- Signaling ↔ Preventing Corruption
PBAS

Goals

Incentives ↔ Measures

Target
Do people agree on goals and can they be measured well?

Which people or units should be held accountable?

What measures can reliably, fairly and robustly reflect attainment of goals?

What targets will focus attention appropriately?

What incentives will motivate changes in behavior?
Performance-Based Accountability Systems Should Not Be Static

PBAS
Goals
Incentives
Measures
Target

Implementation and Monitoring
Refine
Effectiveness