Multiple Measures in the Context of Performance-Based Accountability

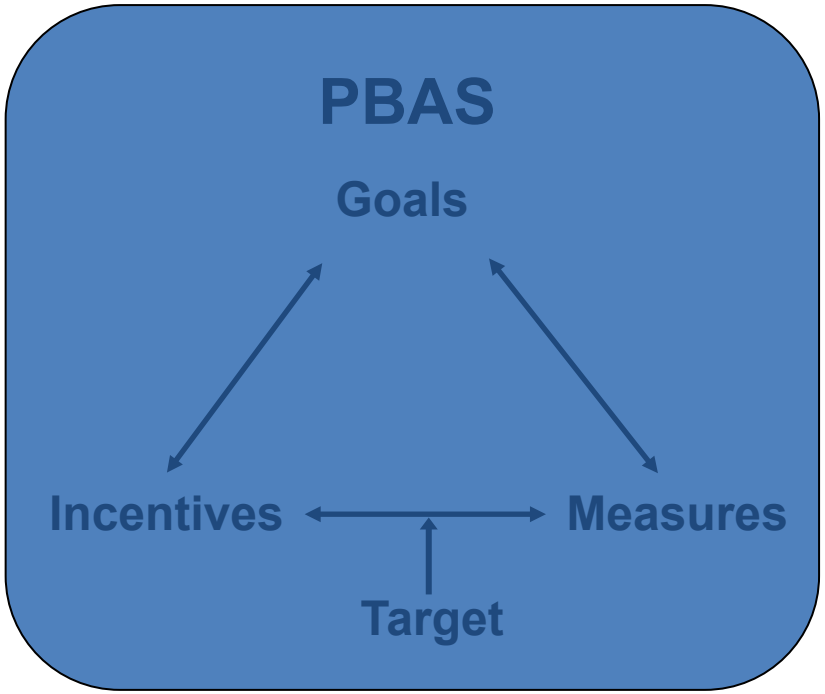
Brian Stecher

Reidy Interactive Lecture Series

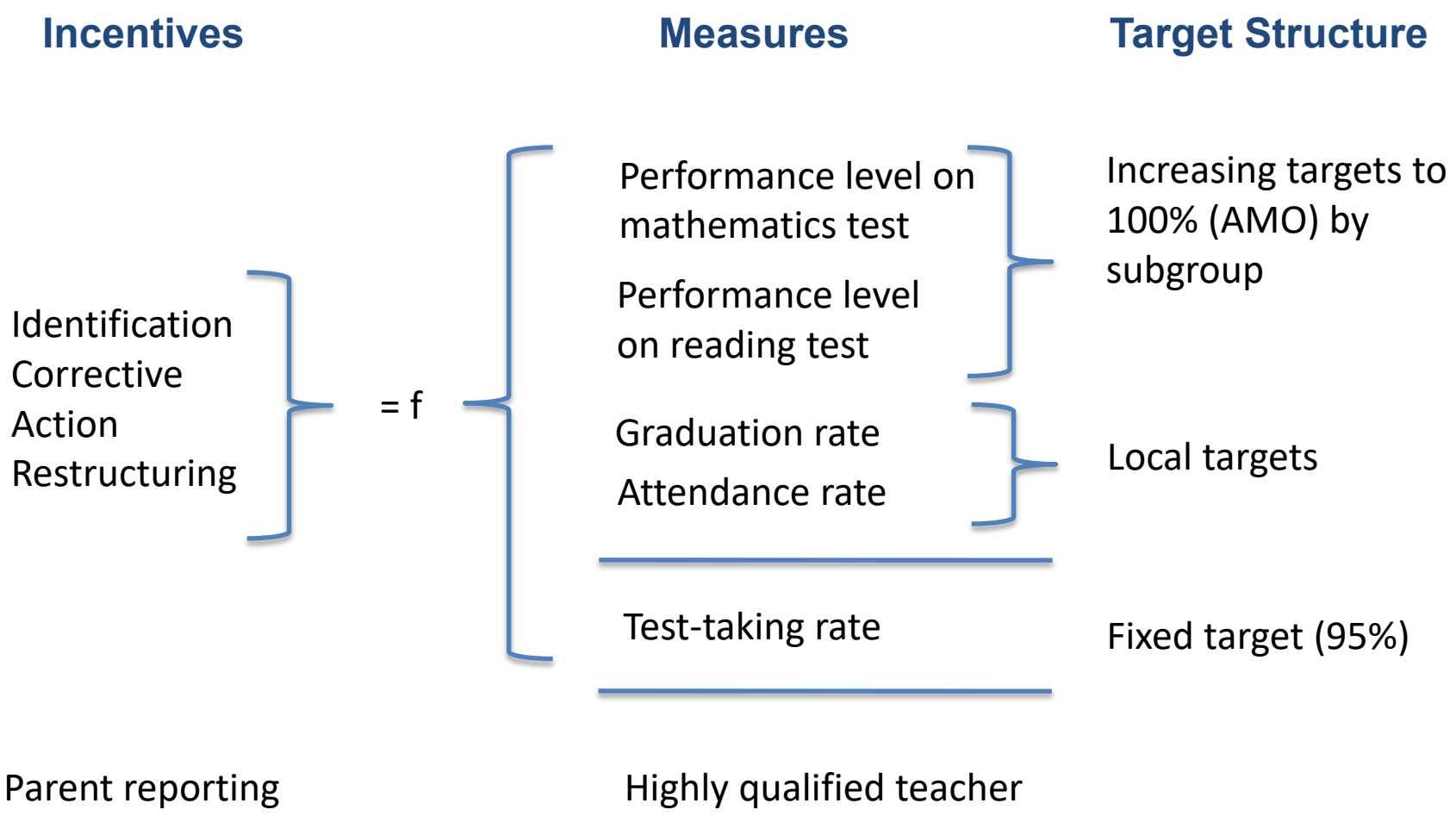
September 23, 2011

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



Twenty States Currently Report Measures Beyond NCLB

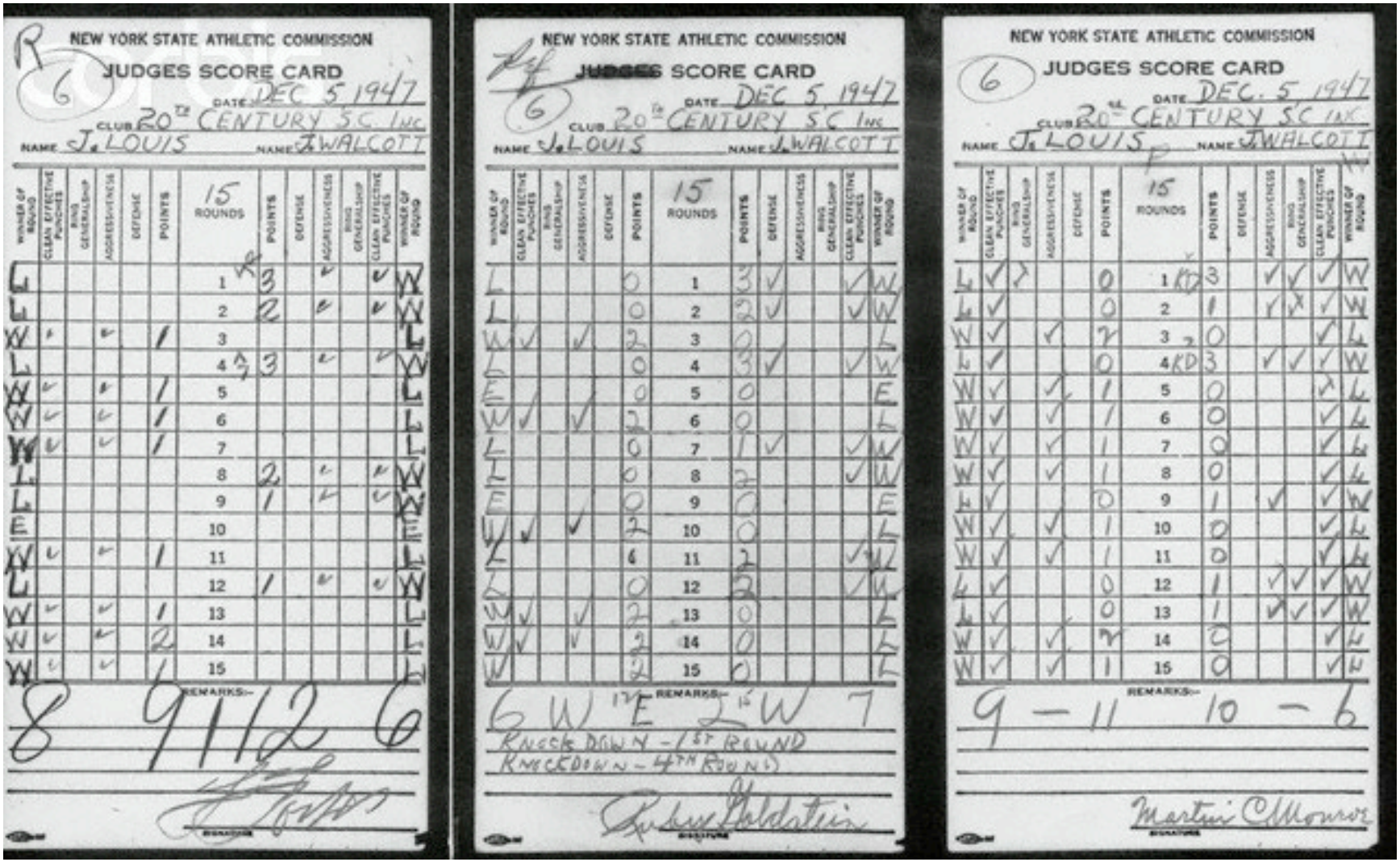
- Outcome indicators included in ratings
  - Additional tested subjects (e.g., history, social studies)
  - College readiness (e.g., ACT, AP)
  - Growth in performance
  - Weighted test scores along whole performance range
- Additional indicators
  - Safe and Supportive Environment
  - Students at-risk 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)



NEW YORK STATE ATHLETIC COMMISSION

JUDGES SCORE CARD

DATE DEC. 5 1947

CLUB 20TH CENTURY S.C. INC.

NAME J. LOUIS

NAME J. WALCOTT

6

15 ROUNDS

	WINNER OF ROUND	CLEAN EFFECTIVE PUNCHES	RING GENERALSHIP	AGGRESSIVENESS	DEFENSE	POINTS
1	L			✓		3
2	L			✓		2
3	W	✓			✓	1
4	L			✓		3
5	W	✓			✓	1
6	W	✓			✓	1
7	W	✓			✓	1
8	L			✓		2
9	L			✓		1
10	E					
11	W	✓			✓	1
12	L			✓		1
13	W	✓			✓	1
14	W	✓			✓	2
15	W	✓			✓	1

REMARKS:

8 9 12 6

NEW YORK STATE ATHLETIC COMMISSION

JUDGES SCORE CARD

DATE DEC 5 1947

CLUB 20TH CENTURY S.C. INC.

NAME J. LOUIS

NAME J. WALCOTT

6

15 ROUNDS

	WINNER OF ROUND	CLEAN EFFECTIVE PUNCHES	RING GENERALSHIP	AGGRESSIVENESS	DEFENSE	POINTS
1	L					0
2	L					0
3	W	✓		✓		2
4	L					0
5	E					0
6	W	✓		✓		2
7	L					0
8	E					0
9	E					0
10	W	✓		✓		2
11	L					1
12	L					0
13	W	✓		✓		2
14	W	✓		✓		2
15	W	✓		✓		2

REMARKS:

6 W 12 E 2 W 7

KNOCK DOWN - (1ST ROUND)

KNOCKDOWN - (4TH ROUND)

NEW YORK STATE ATHLETIC COMMISSION

JUDGES SCORE CARD

DATE DEC. 5 1947

CLUB 20TH CENTURY S.C. INC.

NAME J. LOUIS

NAME J. WALCOTT

6

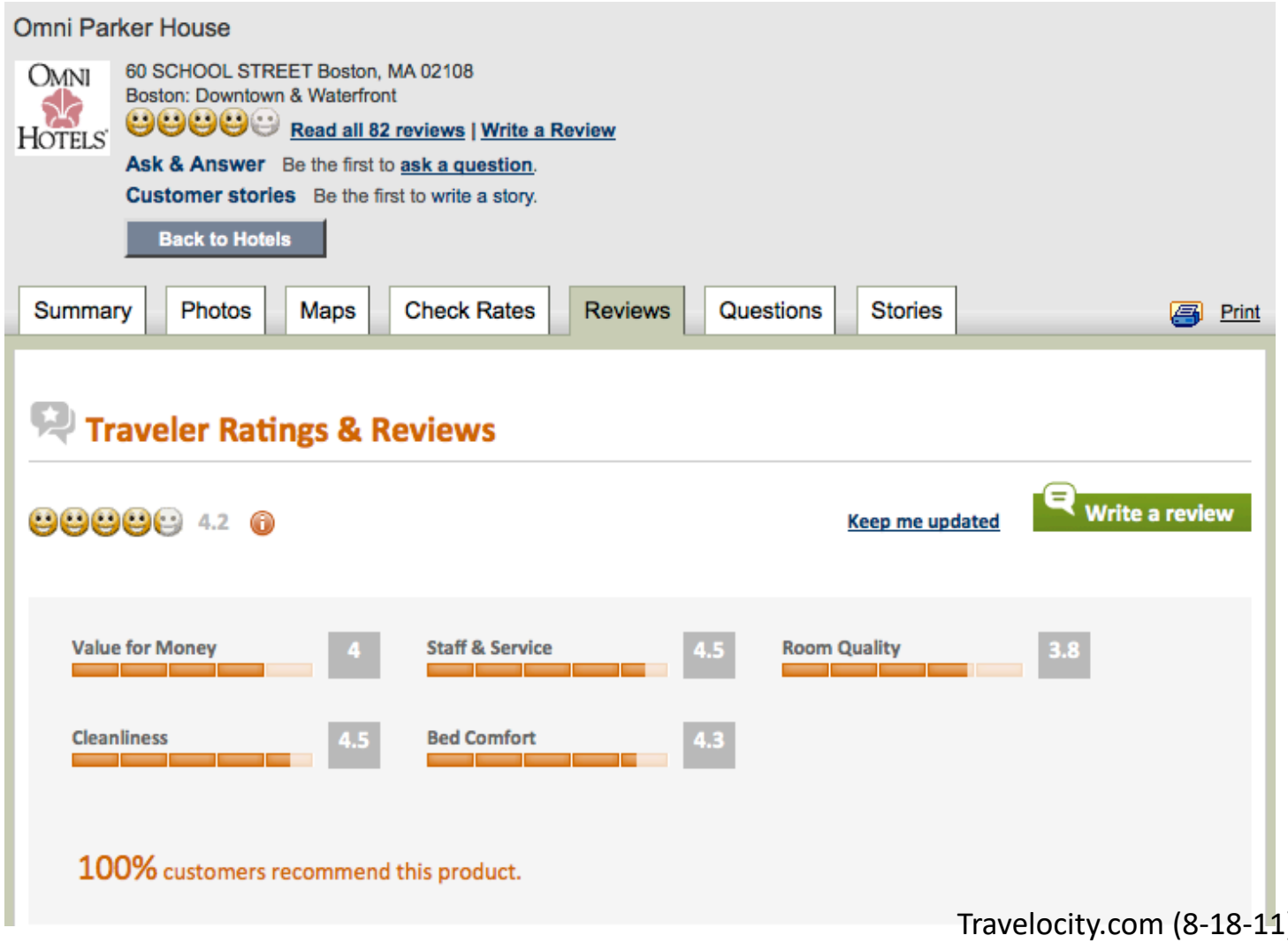
15 ROUNDS

	WINNER OF ROUND	CLEAN EFFECTIVE PUNCHES	RING GENERALSHIP	AGGRESSIVENESS	DEFENSE	POINTS
1 KD	L	✓	✓			0
2	L	✓				0
3	W	✓		✓		2
4 KD	L	✓				0
5	W	✓		✓		1
6	W	✓		✓		1
7	W	✓		✓		1
8	W	✓		✓		1
9	L	✓				0
10	W	✓		✓		1
11	W	✓		✓		1
12	L	✓				0
13	L	✓				0
14	W	✓		✓		2
15	W	✓		✓		1

REMARKS:

9 — 11 10 — 6

Multiple Measures: More Complete Information



Travelocity.com (8-18-11)

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**

TripAdvisor.com (8-18-11)

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

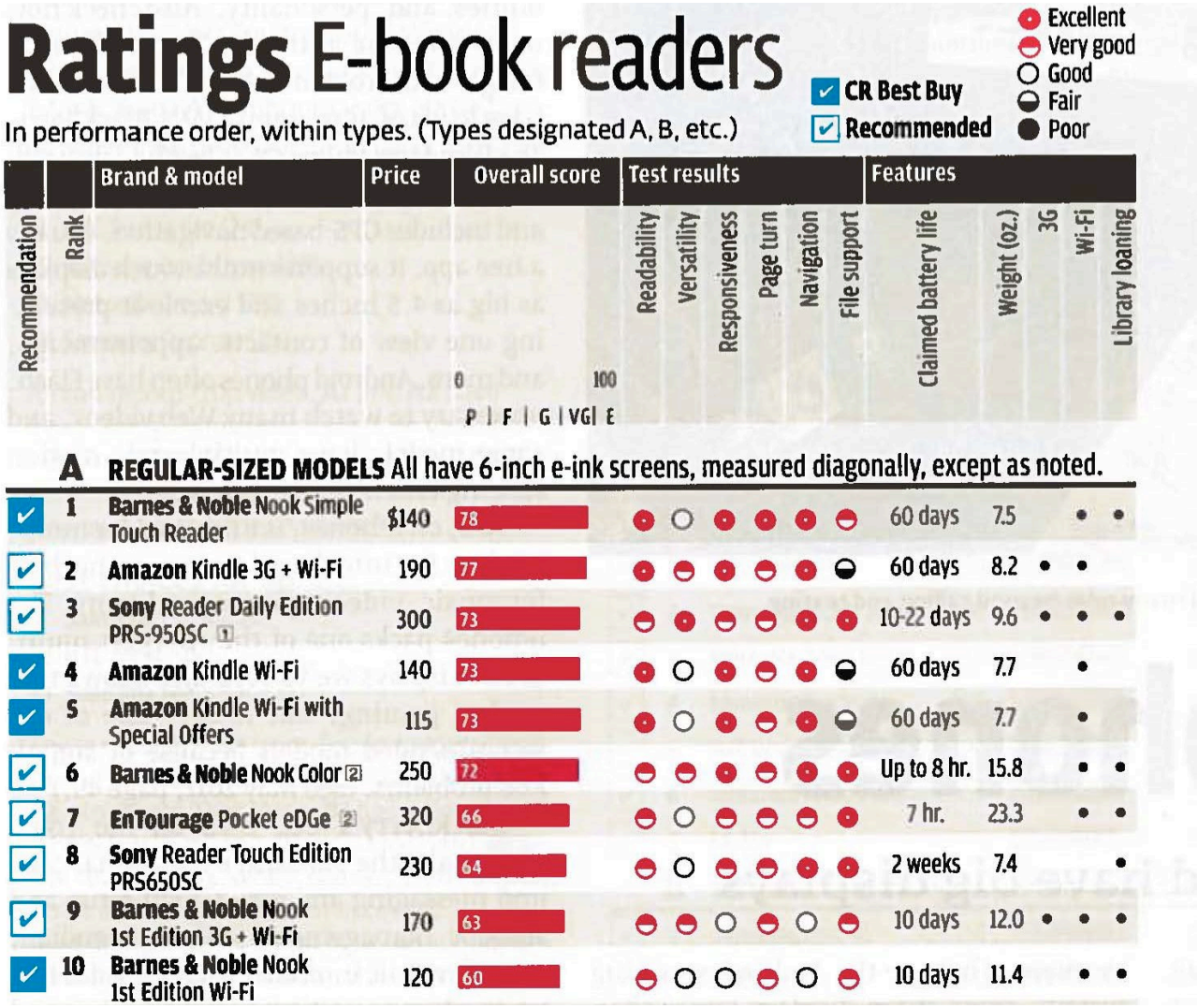
- Academic skills
- Practical and Fine Arts
- Health and Nutrition
- Social Skills
- Psychological Skills
- Citizenship

- 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.)

CR Best Buy
Recommended

Excellent
Very good
Good
Fair
Poor

Recommendation	Rank	Brand & model	Price	Overall score (0 P | F | G | VG | E 100)	Test results: Readability	Versatility	Responsiveness	Page turn	Navigation	File support	Features: Claimed battery life	Weight (oz.)	3G	Wi-Fi	Library loaning
	A	**REGULAR-SIZED MODELS** All have 6-inch e-ink screens, measured diagonally, except as noted.													
✓	1	**Barnes & Noble** Nook Simple Touch Reader	$140	78	●	○	●	●	●	◒	60 days	7.5		•	•
✓	2	**Amazon** Kindle 3G + Wi-Fi	190	77	●	◒	●	◒	●	◒	60 days	8.2	•	•	
✓	3	**Sony** Reader Daily Edition PRS-950SC [1]	300	73	◒	●	◒	◒	●	●	10-22 days	9.6	•	•	•
✓	4	**Amazon** Kindle Wi-Fi	140	73	●	○	●	◒	●	◒	60 days	7.7		•	
✓	5	**Amazon** Kindle Wi-Fi with Special Offers	115	73	●	○	●	◒	●	◒	60 days	7.7		•	
✓	6	**Barnes & Noble** Nook Color [2]	250	72	◒	◒	●	◒	●	●	Up to 8 hr.	15.8		•	•
✓	7	**EnTourage** Pocket eDGe [2]	320	66	◒	○	◒	◒	◒	●	7 hr.	23.3		•	•
✓	8	**Sony** Reader Touch Edition PRS650SC	230	64	◒	○	◒	◒	●	●	2 weeks	7.4			•
✓	9	**Barnes & Noble** Nook 1st Edition 3G + Wi-Fi	170	63	◒	◒	○	◒	○	◒	10 days	12.0	•	•	•
✓	10	**Barnes & Noble** Nook 1st Edition Wi-Fi	120	60	◒	○	○	◒	○	◒	10 days	11.4		•	•

Multiple Measures: More Sensitive Targeting

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

Content Area	2010–11 API Test Weights		
	K–5	6–8	K–8
CST/CMA/CAPA in ELA	56.5%	51.4%	54.2%
CST/CMA/CAPA in Mathematics	37.6%	34.3%	36.1%
CST/CMA/CAPA in Science, Grades 5 and 8	5.9%	7.1%	6.5%
CST in History–Social Science, Grade 8	--	7.1%	3.2%

School Content Area Weights for Grades 9–12

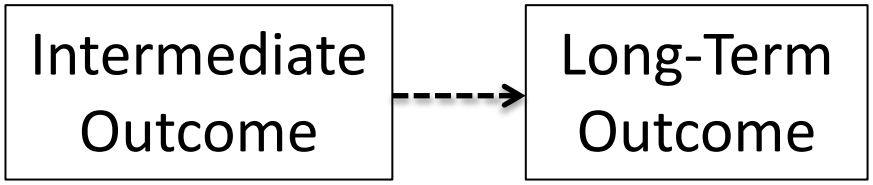
Content Area	2010–11 API Test Weights
	9–12
CST/CMA/CAPA in ELA, Grades 9–11	27.1%
CST/CMA/CAPA in Mathematics, Grades 9–11	18.1%
CST/CMA/CAPA in Science, Grades 9–11	22.9%
CST in History–Social Science, Grades 9–11	13.9%
CAHSEE ELA, Grades 10–12	9.0%
CAHSEE Mathematics, Grades 10–12	9.0%

Challenge:
Measuring Performance that Matters

Long-Term Outcome

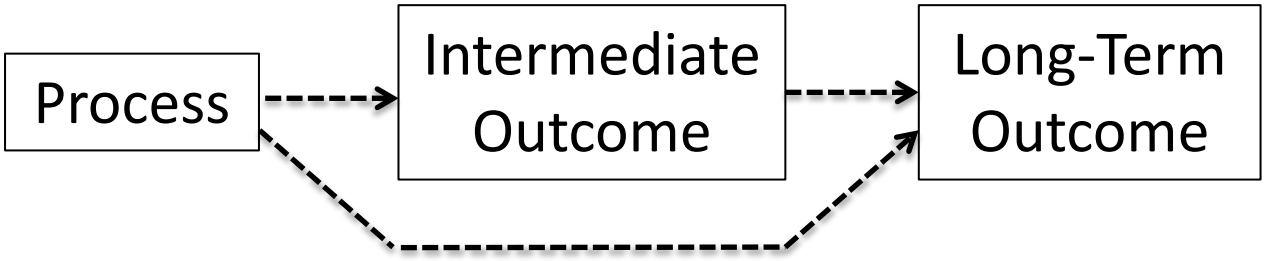
- College Success
- Career Success
- Healthy Living
- Citizenship

Challenge:
Measuring Performance that Matters



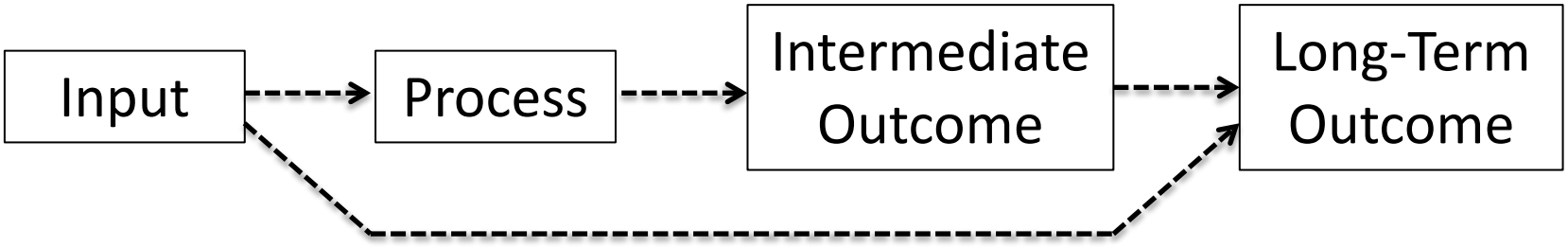
- College Readiness
- Career Readiness
- Course-Taking
- Athletic Participation
- Student Government

Challenge:
Measuring Performance that Matters



- AP Course Offering
- College Counseling
- Art/Music Courses
- Model UN

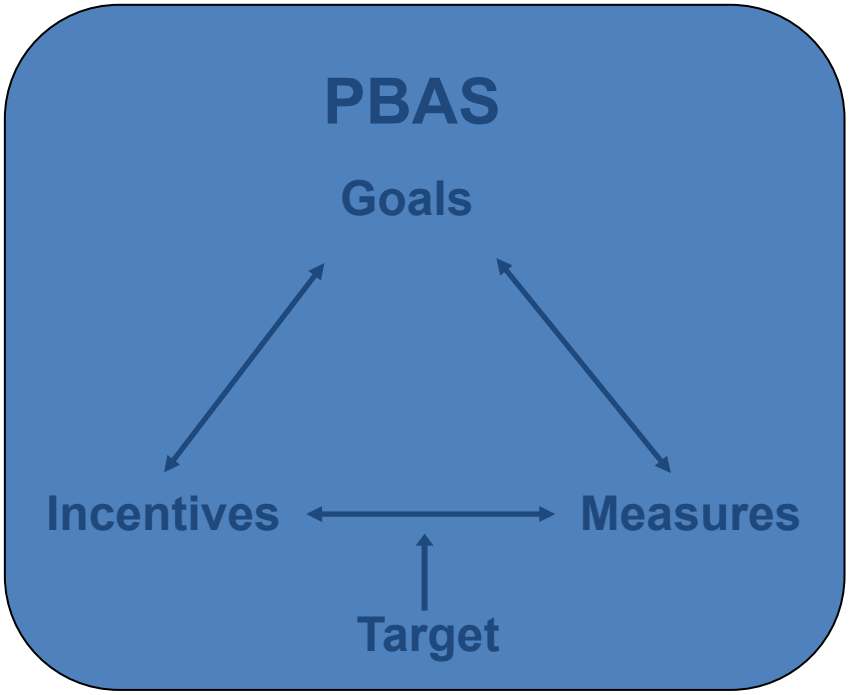
Challenge:
Measuring Performance that Matters

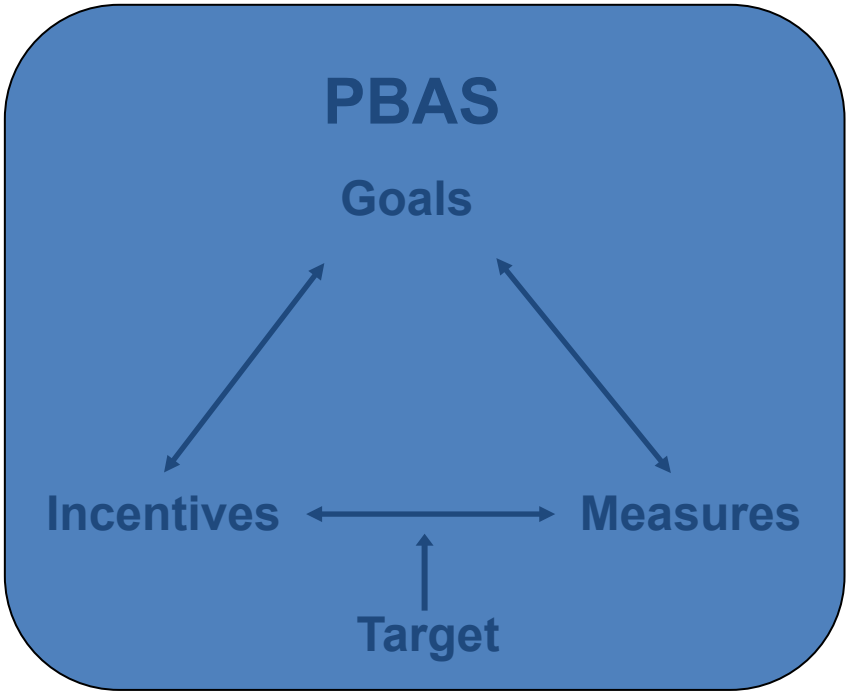


- Qualified Teachers
- Art/Music Facilities
- Community Partners

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?

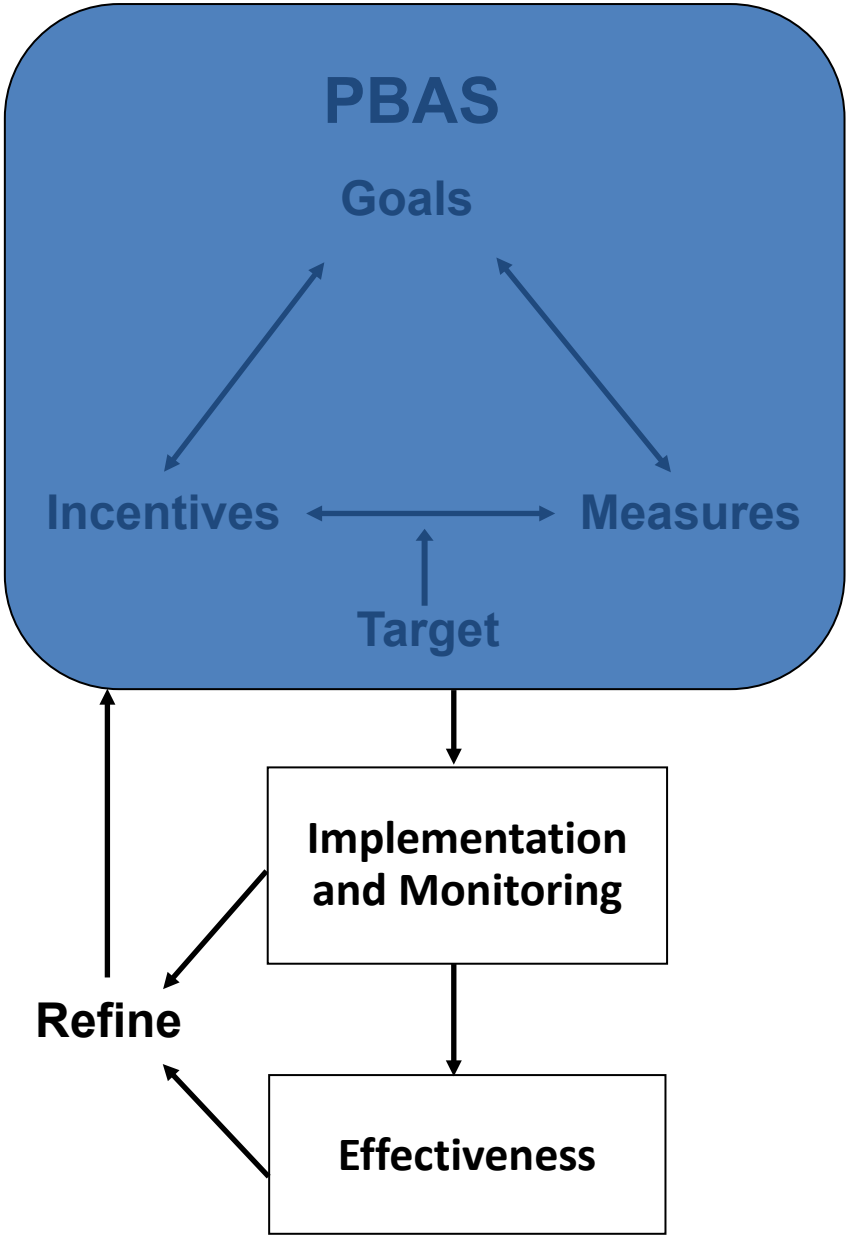
Which people or units should be held accountable?

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



BAYTEX

corbis
5.5
10.0
8.0