What Works?: Exploring State Accountability Indicators to Promote Validity, Equity, and Improvement

Brian Gong

The National Center for the Improvement of Educational Assessment

Presentation to the AERA State Indicators Special Interest Group, Montreal, April 1999

Assessment and accountability systems reflect unique situations.

- Each state is different, because of history, policy players, resources, perceived priorities, sense of “what has already been done” and “what is possible.”
- Assessment and accountability system designs are constructed within these circumstances.

State school accountability systems aggregate data to make decisions and simplify reporting.

- Most states use at least three indicators
  - most widely used for accountability are test scores, student attendance, and dropout; revenue and expenses also reported by majority of states
- Most states report data aggregated at least three levels, some as many as seven

Analysis can help provide validity evidence and focus improvement efforts.

- Each state is unique; needs design that fits state's values and situation.
- Analysis supports efficacy of standards-based educational reform efforts.
- Analysis helps direct improvement efforts, including community support.
- Example: race/ethnicity analysis in three states

Kentucky's assessment system includes multiple measures.

- Two grades each in elementary, middle, and one in high school take state-mandated tests.
- State custom CRTs include seven areas: reading, writing, math, science, social studies, arts & humanities, and practical living/vocational studies. NRT in reading and math also required.
- Tests include multiple choice and constructed response, common and matrix. Writing Portfolios also used. Alternate Portfolio used for alternate assessment.
- Noncognitive measures include attendance, dropout retention, and successful transition to adult life.

Kentucky's school accountability system emphasizes improvement.

- An *index* is constructed annually for each school reflecting performance on CRT and noncognitive measures.
- A *baseline index* averages two years' scores.
- A *growth target* is calculated, based on the baseline index.
- A *growth index* averages the two years' scores subsequent to the baseline.
- *Rewards, sanctions, and targeted assistance* from the state are assigned based on schools' growth compared to their growth targets.

Kentucky's design highlights important policy decisions.

- Assessment policies support instructional policies.
- Effort made to include all students in accountability and "count" each student "equally."
- More subject areas included, longer tests, farther apart, with emphasis on school accountability and comprehensive curriculum modeling, and less on student accountability and testing efficiency.
- Compensatory index used, with a few conjunctive rules for accountability.
- Nominal accountability weights for noncognitive factors low, effective weights even lower.

Kentucky's design highlights important policy decisions (cont.).

- Baselines reset every accountability cycle, up or down, and growth targets set accordingly.
- Accountability school performance judgments trigger (appealable) school consequences, including further school and personnel review. Personnel decisions triggered by separate process.
- Growth targets set to bring all schools (close to) 100 at same time.
- School consequences based only on improvement; no consideration of how high or low absolute performance.

Disaggregating results helps with validity and improvement.

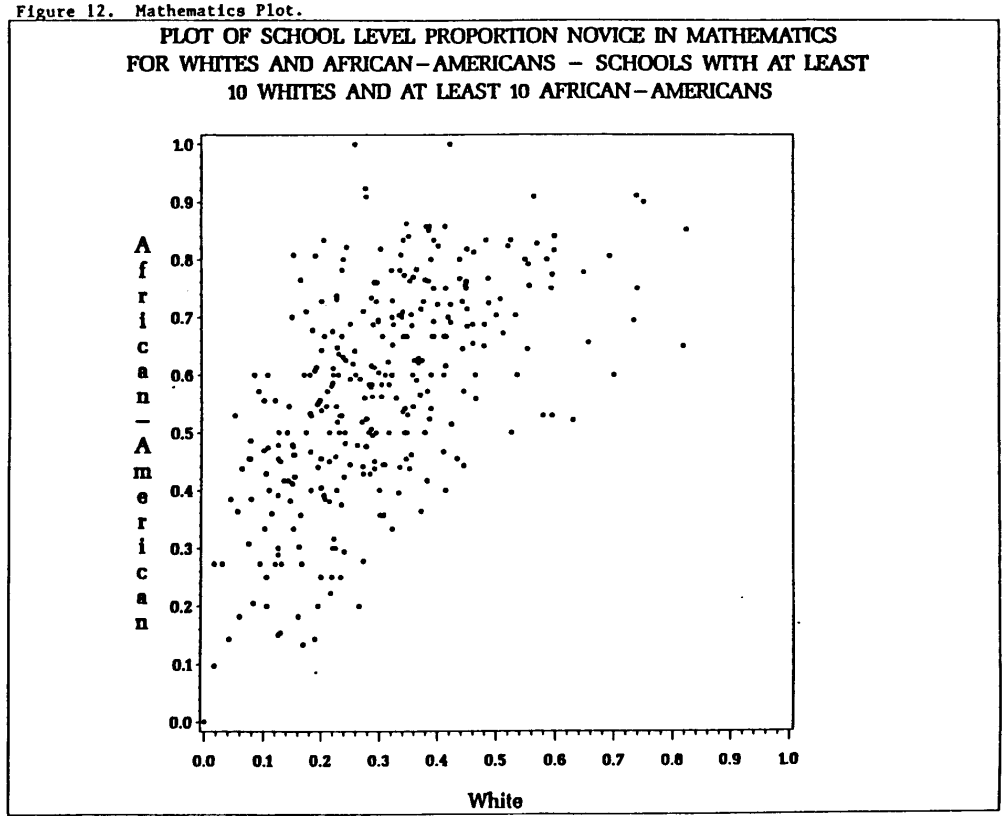
- Three examples around race/ethnicity analyses (different analyses using different indicators)
  - Kentucky: school-level equity
  - Louisiana: race/ethnicity and economic class
  - Massachusetts: race/ethnicity, course-taking patterns

Kentucky: Some equity results

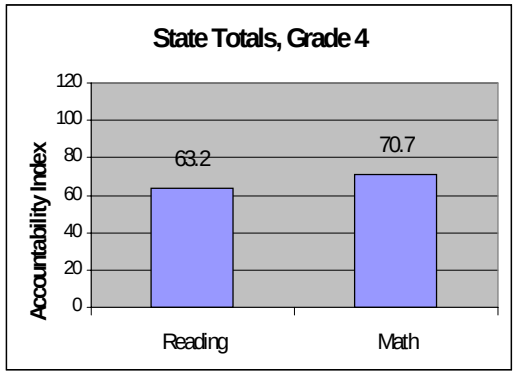
- It was possible for schools to meet improvement goals
- Meeting improvement goals was *not* a function of poverty, race, geography, or “how low a school started”
- Improvements in content areas reflected teacher familiarity, training
- Ethnicity achievement gap varied widely between schools

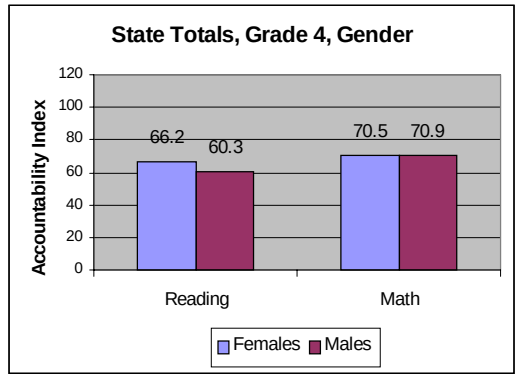
Kentucky's accountability results not related to race, but consider subgroup performance by school.

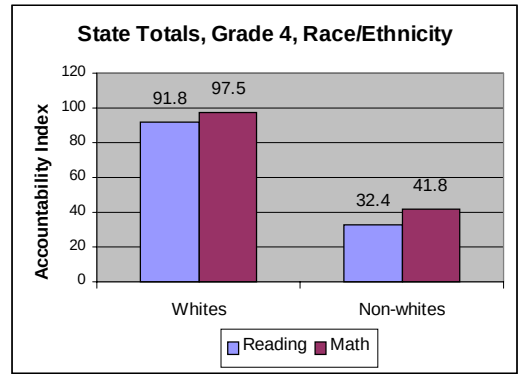
Figure 12. Mathematics Plot.

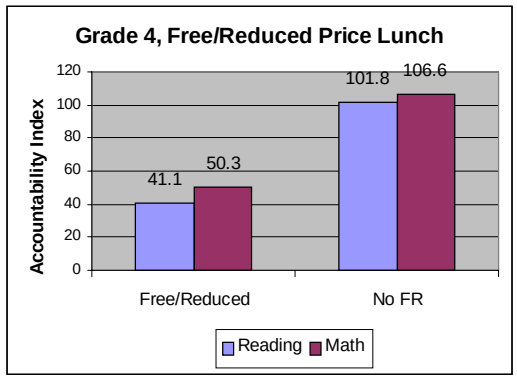


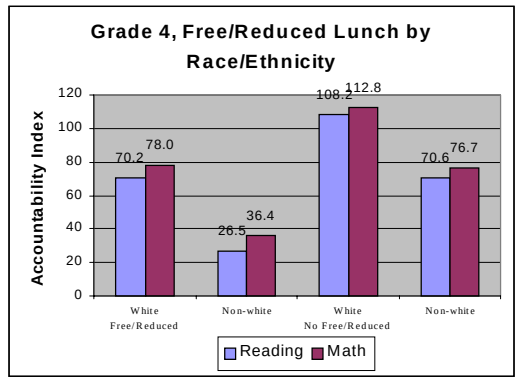
Analyzing multiple indicators may be informative.











To improve, states (and districts/ schools) should consider additional indicators outside of accountability.

Percent Currently Enrolled in "Beginning" and "On-Sequence" Courses, Science						
	Asian/ Pacific Islander	African American	Hispanic/ Latino	Multiple	Native American	White
Not enrolled[1]	4	38	44	40	29	26
Beginning[2]	21	28	30	30	37	16
On-sequence[3]	58	28	20	30	17	53

[1]Student did not report being currently enrolled in any of the science courses identified in the student questionnaire
[2]Integrated Science or Biology course
[3]Chemistry or Physics course

MCAS performance	Commercial test control	Other factors	R-square	Change in R-square
Sci. & Tech.	MAT7 Science		.55	
Sci. & Tech.	MAT7 Science	Biology	.60	.04
Sci. & Tech.	MAT7 Science	Biology, Geometry	.61	.05
Sci. & Tech.	MAT7 Science	Race/Ethnicity	.56	.01
Sci. & Tech.	MAT7 Science	Biology, Geometry, Race/Ethnicity	.61	.00

Percentage of students who reported having completed a biology course						
MCAS Science Level	Asian/ Pacific Islander	African American	Hispanic/ Latino	Multiple	Native American	White
Failing	19	22	45	25	-	8
Proficient	8	22	9	13	-	34