

# Considerations When Selecting and Combining Measures

---

**Robert L. Linn**

*CRESST, University of Colorado at Boulder*

*Paper presented at a Workshop on Multiple Measures sponsored by the Committee on Incentives and Test-Based Accountability. Washington, DC: The National Academies, November 16, 2007*

# Multiple Measures

- Mean different things
- Two meanings that I will not talk about
  1. Let test takers repeat a test (e.g., when they fail a first attempt a graduation test)
  2. NCLB uses of assessments of reading and mathematics as well as participation and another indicator such as attendance
- The first is just a repeat of a single measure not multiple measures and the second falls outside focus because the measures are used in a conjunctive way in accountability

# Confusion about Multiple Measures

- Led a group of journal editors working with the National Council on Measurement in Education to propose using “multiple sources of evidence” rather than “multiple measures.”
- (Letter to Senator Kennedy and Representative Miller, chairs of the relevant Senate and House Committees sent by Rudner, Brookhart, Geisinger, Gitomer, & Fan, April 11, 2007)

# Test Standards

- Clearly indicate that important educational decisions should not be made on the bases of scores on a single test
- Adding information from other sources can enhance the validity and reliability of decisions

# Conjunctive Decision Models

- Reduce reliability and validity of decisions because the overall reliability and validity depends on the reliability and validity of the weakest element in the model
- The conjunctive model used to determine AYP results in many ways for schools to fail to make AYP, but only one way to make it – exceed the target on each measure for every subgroup for which disaggregated reporting is required

# Compensatory Decision Models

- Compensatory composite scores generally have higher reliability and validity than the individual components that make up the composite
- Combining information from multiple measures for decision making, e.g., determining a report card for a school does not preclude the use of the individual pieces of information provided by each measure

# Multiple Measures (Continued)

- Consider two meanings
  1. Multiple ways of assessing students in a given content area (e.g., multiple choice tests and performance assessments, or standardized tests and teacher judgments)
  2. Assessing student achievement in multiple subjects (e.g., reading, writing, mathematics, science) and combining scores in a compensatory system

# Multiple Measures Within a Single Content Area

- Proposal under consideration by Representative Miller for NCLB reauthorization
- Main potential benefit would come from the use of different formats, e.g., constructed response items on district assessment
- Could reduce problems of inflated test scores due to narrow teaching to a single test
- Combination rules would need to be worked out
- Would likely require the setting of new targets for AYP

# Composites of Multiple Measures Across Different Content Areas

- Recognizes the importance of content areas other than reading and mathematics
- Allows performance above expectations in one content area to compensate for somewhat lower performance in another area
- Several states use some form of composite across content areas

# Examples of Multiple Content Area Composites from Four States

- California
- Florida
- Kentucky
- Ohio

# California's Academic Performance Index (API)

- API is a weighted composite of performance on California assessments and improvement on the measures
- API includes performance for “significant” subgroups of at least 100 or more students as well as all tested students in the school

# API Weights for Content Areas Grades 2 through 8

CST English-Language Arts, grades 2-8	0.48
CST Mathematics, grades 2-8	0.32
CST Science, grade 5	0.20
CST History-Social Studies, grade 8	0.20
CAT/6 Survey in Reading, grades 3 & 7	0.06
CAT/6 Survey in Language, grades 3 & 7	0.06
CAT/6 Survey in Spelling, grades 3 & 7	0.03
CAT Survey in Mathematics, grades 3 & 7	0.03
CST in Science, grade 8	0.20
Assignment of 200 in Mathematics, grade 8	0.10

# API Performance Level Scores

Performance Level	CAT/6 Performance	CAHSEE Score	API Score Weight
Advanced	80-99 <sup>th</sup> NPR	Pass	1000
Proficient	60-79 <sup>th</sup> NPR	N/A	875
Basic	40-59 <sup>th</sup> NPR	N/A	700
Below Basic	20-39 <sup>th</sup> NPR	N/A	500
Far Below Basic	1-19 <sup>th</sup> NPR	No Pass	200

# API Target and Reporting

- School Target 800
- Report by API Decile for all schools
- Report by API Decile for similar schools on School Characteristics Index (SCI)

# Florida's Grading of Public Schools

- Schools are awarded grades of A through F based on a composite derived from performance on the Florida Comprehensive Assessment Test (FCAT) and gains in performance on the FCAT

# Florida Grades

<b>A</b>	525 points or more Adequate progress of lowest students in reading and math Test at least 95% of eligible students
<b>B</b>	495 points or more Adequate progress of lowest students in reading and math with two years. Test at least 90% of eligible students
<b>C</b>	435 points or more Adequate progress of lowest students in reading and math with two years. Test at least 90% of eligible students
<b>D</b>	395 points or more Test at least 90% of eligible students
<b>F</b>	Fewer than 395 points OR Less than 90% of eligible students tested

# Florida Points

- Schools earn one point for each percent of students who score at levels 3, 4, or 5 in (a) **reading**, (b) **math**, and (c) **science**
- Schools earn one point for each percent of students who score 3.5 or higher on the **writing** exam
- Schools earn one point for each percent of students who make **learning gains** in (a) **reading** and (b) **math**

# Kentucky's Accountability System

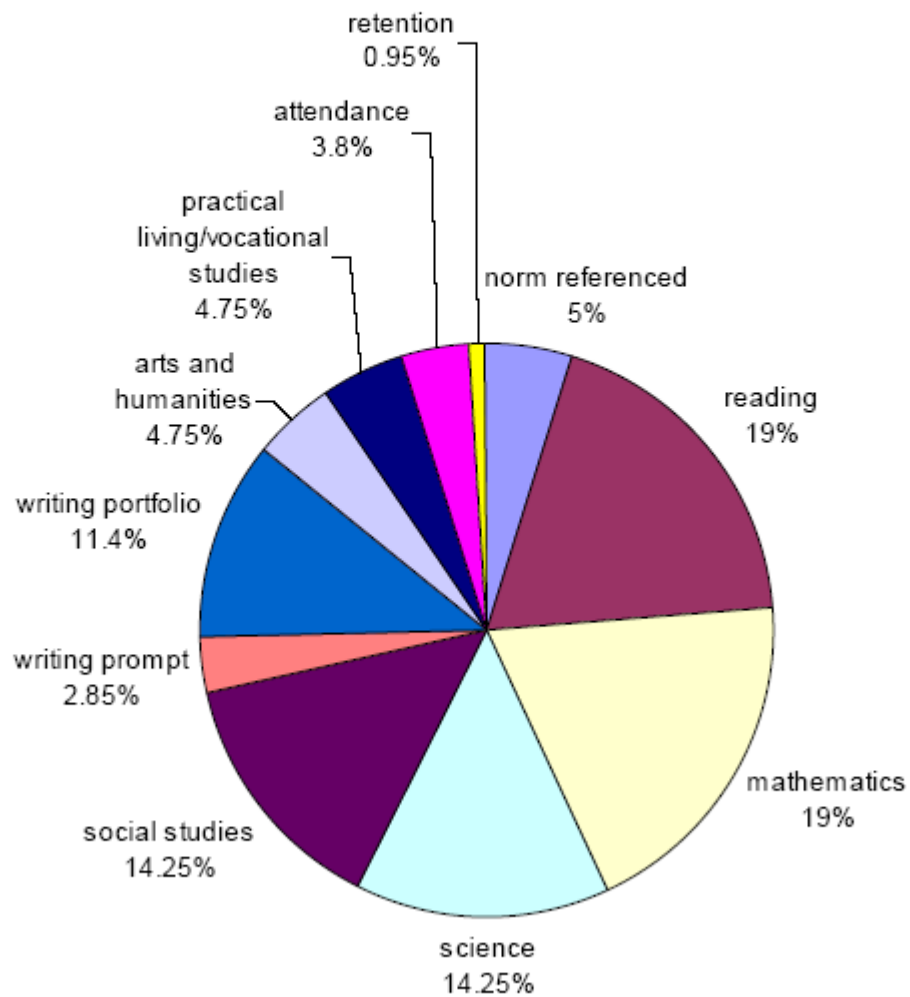
- Uses an academic index that is a weighted composite of performance on Kentucky assessments in 7 content areas plus scores on a norm-referenced test
- Accountability index is a composite of the academic index and non-academic indicators

	KCCT						WRITING*		Norm-Referenced Assessments	Readiness Examination
Grade	Reading	Mathematics	Science	Social Studies	Arts & Humanities	Practical Living/ Voc. Studies	On-Demand	Writing Portfolio	Reading / Language Arts and Mathematics	High School and College Diagnostic/Predictive Measurement
End of Primary (grade 3)	MC, OR	MC, OR							Optional ***	
4	MC, OR	MC, OR	MC, OR			MC		Analytic		
5	MC, OR	MC, OR		MC, OR	MC, OR		Analytic One prompt and MC			
6	MC, OR	MC, OR							MC (Fall)	
7	MC, OR	MC, OR	MC, OR			MC		Analytic		
8	MC, OR	MC, OR		MC, OR	MC, OR	ILP**	Analytic One prompt and MC			EXPLORE <i>English, Reading, Mathematics, Science</i> (Fall)
9						ILP**				
10	MC, OR					MC, ILP**				PLAN <i>English, Reading, Mathematics, Science</i> (Fall)
11		MC, OR	MC, OR	MC, OR	MC, OR	ILP**				
12						ILP**	Analytic Two prompts (Fall)	Analytic (Spring)		

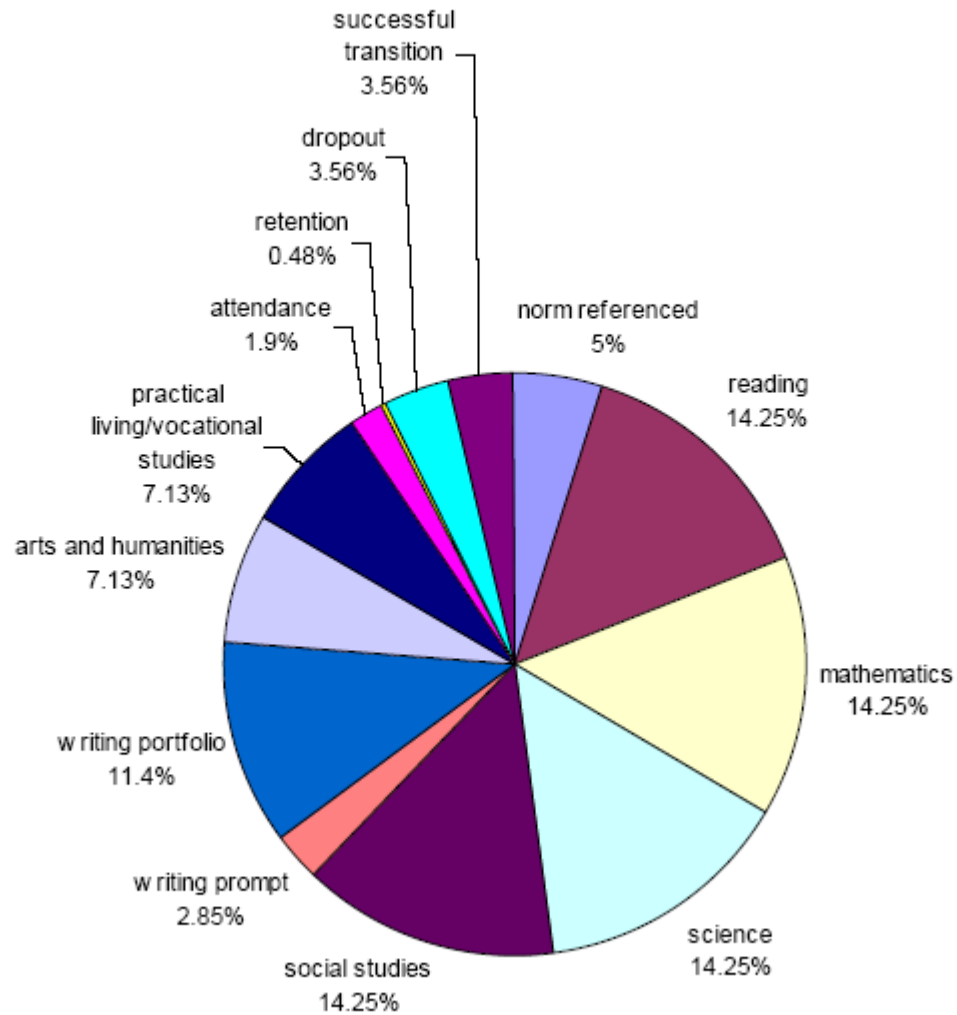
# Weights for Performance Levels

Level	Category	Points
Novice	Low	0
	Medium	13
	High	26
Apprentice	Low	40
	Medium	60
	High	80
Proficient		100
Distinguished		140
Goal for schools by 2014		100

COMMONWEALTH ACCOUNTABILITY TESTING SYSTEM  
ELEMENTARY SCHOOL WEIGHTS



# COMMONWEALTH ACCOUNTABILITY TESTING SYSTEM HIGH SCHOOL WEIGHTS



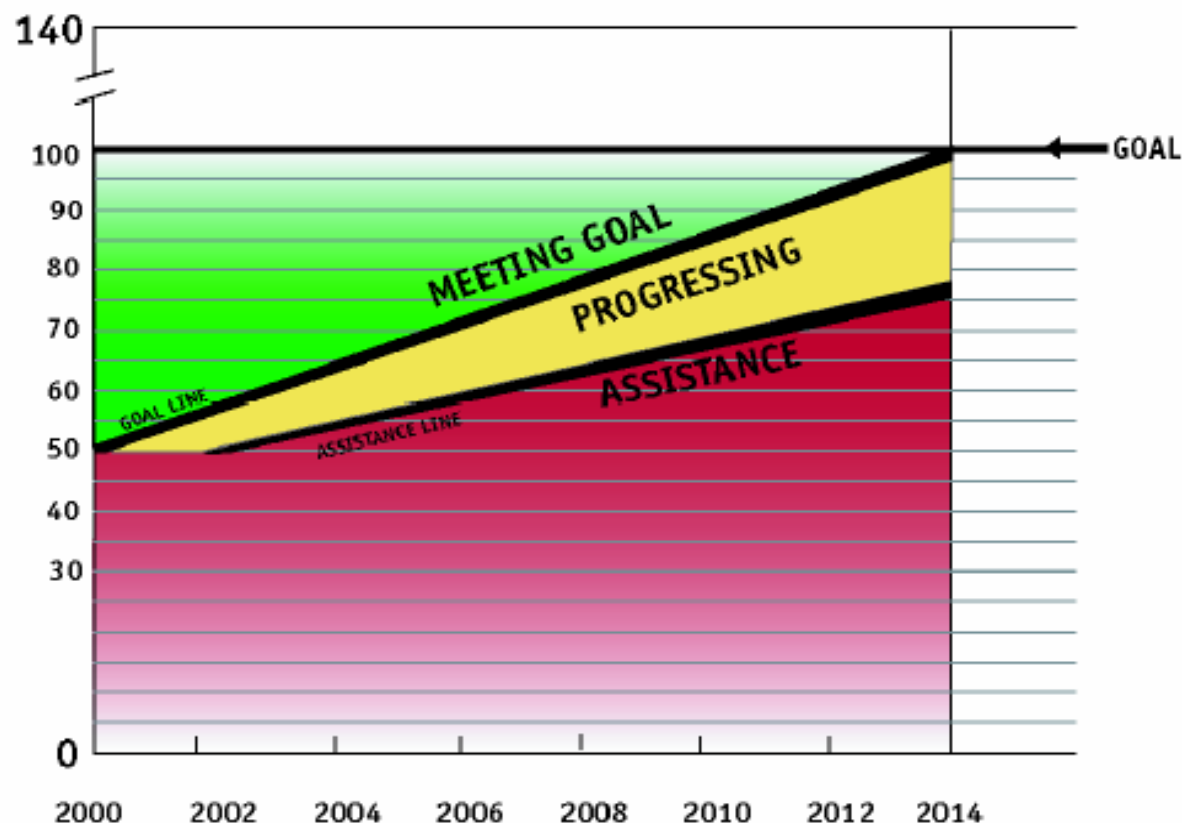
# Commonwealth Accountability Testing System Long-term Accountability Model

Each school will have a customized chart like this one based on its own baseline scores, and a goal of 100 by year 2014. This example is for a hypothetical school with a baseline of 50.

**MEETING GOAL:** Schools at or above the Goal Line will receive three shares of rewards (per full-time equivalent certified staff member). The Goal Line includes a fairness margin. If a school is that close to its Goal Line, it will still receive three shares of rewards.

**PROGRESSING:** Schools scoring below their Goal Line but at or above their Assistance Line will be classified as "Progressing". A school in this zone receives 1/2 share of rewards (per full-time equivalent certified staff member) provided it obtains an index greater than it did in the previous biennium. Schools with stable or declining scores will receive no rewards but also no consequences.

**ASSISTANCE:** Schools falling below the Assistance Line may be eligible for a "scholastic audit" to determine how much assistance or help is needed and whether it should be monetary (Commonwealth School Improvement Funds) and/or human resources (highly-skilled educators).



**SCHOOL RECOGNITION POINTS:** When a school passes a RECOGNITION POINT based on its total accountability score, it will receive a share of rewards as a one-time recognition of achievement. Five levels of recognition were established by the Kentucky Board of Education. The five recognition points are located at accountability indices of 55, 68, 77, 88 and 100.

**PACE SETTER SCHOOLS:** The top 5% of schools will be designated as "Pace Setters" if they have met or exceeded the fourth recognition point. They will receive 1 share of rewards provided that they are not receiving rewards under any other provision and provided that they have not declined in both the previous two biennia.

**Note:** To receive rewards, a school must also meet both the dropout reduction and novice reduction requirements.

# Ohio's School and District Report Cards

- Schools and districts are given one of five ratings (Excellent, Effective, Continuous Improvement, Academic Watch, or Academic Emergency) based on a combination of four components (1) the proportion of state indicators they meet, (2) AYP status, (3) a Performance Index Score, and (4) Growth on the Performance Index (to be replaced by value-added in 2008)

# State Indicators

- The state indicators vary by year, but are based on the number of assessments given at all tested grades.
- To earn each indicator, a school or district needs to have a specified percentage of students score at the proficient level or higher on a given assessment.

<b>2006-07 State Indicators</b>		
<b>Grade</b>	<b>Tests</b>	<b>Proficient or Above Standard</b>
3	Reading and Math Achievement	75%
4	Reading, Writing and Math Achievement	75%
5	Reading, Math, Social Studies, and Science Achievement	75%
6	Reading and Math Achievement	75%
7	Reading, Writing, and Math Achievement	75%
8	Reading, Math, Social Studies, and Science Achievement	75%
Ohio Graduation Test- 10 <sup>th</sup> grade	Reading, Writing, Math, Social Studies, and Science	75%
Ohio Graduation Test- 11 <sup>th</sup> grade	Reading, Writing, Math, Social Studies, and Science	85%
<b>Non-test Indicators</b>		
	Graduation Rate	90%
	Attendance Rate	93%
<b>30 Total State Indicators</b>		

## Performance Index

The Performance Index (PI) is a calculation that measures achievement/OGT test performance at the 3<sup>rd</sup>, 4<sup>th</sup>, 5<sup>th</sup>, 6<sup>th</sup>, 7<sup>th</sup>, 8<sup>th</sup>, and 10<sup>th</sup> (OGT) grade levels based upon the number of students at each performance level. The PI is calculated by assigning a weighted score to each performance level in the following manner:

- Untested students - 0 points
- Below Basic/Limited - 0.3 points
- Basic - 0.6 points
- Proficient - 1.0 points
- Accelerated – 1.1 points
- Advanced - 1.2

The percentage of students at each performance level is then multiplied by their respective weight, and the totals for each performance level are summed to get the building's overall Performance Index score.

# 2006-07 Ratings Table

	Indicators Met (District #'s in parenthesis)		Performance Index Score		Change in PI Score		AYP Status
Excellent	94% - 100% (29 or 30)	or	100 to 120			and	Met/Not Met**
Effective	75% - 93.9% (23 to 28)	or	90 to 99.9			and	Met/Not Met**
Continuous Imprvmnt.	0% - 74.9% (0 to 22)	or	0 to 89.9		PI Score Improvement Criteria Met*	and	Met
	50% - 74.9% (15 to 22)	or	80 to 89.9	or		and	Not Met
Academic Watch	31% - 49.9% (10 to 14)	or	70 to 79.9	or		and	Not Met
Academic Emergency	0% - 30.9% (9 or fewer)	and	0 to 69.9			and	Not Met

\*10 point increase over 2 years with at least 3 points in the most recent year

\*\*Schools and Districts that miss AYP for three consecutive years in more than one student group in the most recent year can be rated no higher than Continuous Improvement.

# Summary

- As illustrated by the four state examples, scores from multiple measures can be combined in a variety of ways
- Creating a weighted has advantages over the conjunctive treatment of results from multiple measures
- Weights can be used to identify relative priorities

# Summary (Continued)

- Composite scores can enhance validity and reliability of decisions
- Combining multiple measures into composites allows for the addition of new measures without creating additional hurdles to be cleared to show satisfactory performance
- A conjunctive approach is no stronger than the weakest measure and the likelihood of failure and of false negatives increases as the number of measures increases