

# Making Valid Inferences About School Effectiveness from Multiple Measures

*National Academy of Sciences, Center for Education, Board on Testing and Assessment  
Workshop on Multiple Measures*

*Mitchell D. Chester*

*Senior Associate Superintendent*

*November 16, 2007*

# Concerns about Accountability Systems & High Stakes Testing

- Narrowing of instructional content & services
- Restricted teacher authority & professionalism
- Some student groups penalized disproportionately
- **Narrow notions of “effectiveness”**
- **Inaccurate inferences**
  - Score volatility & decision inconsistency
  - Over-identification of schools

# The Remedy: Multiple Measures (Multiple Sources of Evidence) **BUT ...**

- Use of multiple measures does not automatically improve:
  - Validity of conclusions
  - Reliability of conclusions
- Method for combining measures is critical consideration
- Policy goals should drive use

# Types of Multiple Measures

- Measures of different constructs
- Different measures of the same construct
- Multiple opportunities

# Combining Measures

- **Conjunctive**
  - Attainment of standards on multiple measures required
- **Compensatory**
  - Stronger results on some measures compensate for weaker results on others
- **Confirmatory**
  - One measure used to validate another
- **Complementary**
  - Attainment of standard on one of multiple measures suffices

# Policy Context Examples

- **Minimize possibility of false negatives**
  - **Compensatory** approach to combining measures of same construct
  - **Complementary** approach to combining multiple opportunities
- **Maximize certitude of performance judgments**
  - **Confirmatory** approach to combining measures
- **Privilege multiple effectiveness constructs**
  - **Complementary** approach to combining measures of different constructs
- **Ensure mastery of multiple constructs**
  - **Conjunctive** approach to combining measures of different constructs
- **Identify potential concerns**
  - **Conjunctive** approach to combining measures

# Multiple Measures & AYP

	<b>Conjunctive (AND)</b>	<b>Confirmatory (AND)</b>	<b>Compensatory (+/-)</b>	<b>Complementary (OR)</b>
<b>Measures of Different Constructs</b>	Must meet proficiency targets in reading and math, participation targets in reading and math, graduation and attendance rate targets for each group to meet AYP	For districts to enter "improvement" status, must miss AYP in each grade band (elementary, middle, and high school)		Improvement of non-proficient groups fulfills AYP proficiency requirement (safe harbor and growth models)
<b>Different Measures of the Same Constructs</b>			Assessment results averaged across grades so that stronger proficiency rates in a subject at one grade may compensate for weaker rates at another grade	Performance of students with disabilities on either standard or alternative assessment applied to proficiency rate
<b>Multiple Opportunities</b>		Must miss AYP in same subject for multiple years to enter improvement status; By maintaining the same annual measurable objectives for two or three years, must miss proficiency targets for each of the years to miss AYP consecutively	Stronger of two-year average or most recent year's proficiency rate used to determine AYP	

# Ohio Classification System

	Performance Indicators		Performance Index Score		Growth Calculation		NCLB AYP Requirement
<b>Excellent</b>	94% to 100%	<i>or</i>	100 to 120		<p><b>Value Added</b>            Beginning in 2007-08, Ohio is incorporating a measure of individual student grade-grade achievement gains to help determine school building and district classifications <b>2+ years of above average</b> gains improves rating <b>3+ years of below average</b> gains reduces rating</p>	<i>and</i>	Met / Missed* AYP
<b>Effective</b>	75% to 93%	<i>or</i>	90 to 99			<i>and</i>	Met / Missed* AYP
<b>Continuous Improvement</b>	0% to 74%	<i>and</i>	0 to 89			<i>and</i>	Met AYP
	50% to 74%	<i>or</i>	80 to 89	<i>or</i>		<i>and</i>	Missed AYP
<b>Academic Watch</b>	31% to 49%	<i>or</i>	70 to 79	<i>or</i>		<i>and</i>	Missed AYP
<b>Academic Emergency</b>	0% to 31%	<i>and</i>	0 to 69		<i>and</i>	Missed AYP	

\* Can miss AYP and earn "Excellent" or "Effective" designation for up to two years -- third year missing AYP, classification drops.

# Multiple Measures & Ratings

	<b>Conjunctive (AND)</b>	<b>Confirmatory (AND)</b>	<b>Compensatory (+/-)</b>	<b>Complementary (OR)</b>
<b>Measures of Different Constructs</b>	Will remain classified Excellent or Effective if meet performance indicator or performance index standard AND meet AYP requirements		Performance index score is composite of assessment scores from all grades and subjects	For lowest two categories, higher classifications driven by status or by strong two-year gains; Meeting AYP results in Continuous Improvement classification when otherwise Academic Emergency or Academic Watch
<b>Different Measures of the Same Constructs</b>		Classified Academic Emergency only if miss benchmarks for each of four measures		Classification driven by higher of performance indicator status and performance index score; strong or weak value-added gains drive higher or lower ratings
<b>Multiple Opportunities</b>		For highest two classifications, will drop to Continuous Improvement if miss AYP more than two consecutive years; strong or weak value-added gains drive higher or lower ratings if present for multiple years		

# Three Propositions

1. Expanding sources of evidence for school & district effectiveness could:
  - Be powerful incentive for school improvement and focus on all students
  - Mask:
    - ✓ Low outcomes for some students
    - ✓ Low results in some areas
2. How the sources of evidence are combined is critical
3. Policy objectives should drive decisions

# To extent that future ESEA policy goals include:

## Policy Goal

- Focus on historically underserved students

## Implication

- Avoid rules for combining measures that allow low group performance to be masked

# To extent that future ESEA policy goals include:

## Policy Goal

- Ensure that all students learn to read *and* to do math

## Implication

- Adopt rules for combining measures that ensure mastery of both subjects

# To extent that future ESEA policy goals include:

## Policy Goal

- Promote measures of achievement beyond those of reading and math

## Implication

- Adopt rules for combining measures that signal the policy intent (e.g., Is it OK for some students not to learn to read and do math?)

# To extent that future ESEA policy goals include:

## Policy Goal

- Focus on all students reaching a level of proficiency

## Implication

- Adopt rules for combining measures that increase the certitude of performance judgments

# To extent that future ESEA policy goals include:

## Policy Goal

- Privilege teacher assessments of student achievement

## Implication

- Adopt rules for combining teacher and standardized measures of student achievement

# To extent that future ESEA policy goals include:

## Policy Goal

- Identify low performing schools

## Implication

- Adopt rules for combining measures that identify potential concerns

# To extent that future ESEA policy goals include:

## Policy Goal

- Identify the **lowest** performing schools

## Implication

- Adopt rules for combining measures that maximize the certitude of performance judgments & minimize false negatives