

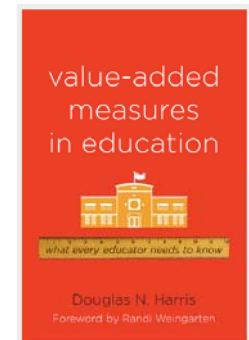
# Value-Added Measures in Education: Part II

Douglas N. Harris

Associate Professor of Educational Policy and Public Affairs  
University of Wisconsin at Madison

April 29, 2011

State of Missouri



# Chapters 1-4: Arguments for VA


1. Exploring the Potential of Value-Added Performance Measures
2. Using—and Misusing—Achievement Tests
3. Measuring Student Growth
4. Creating Value-Added Measures
5. Understanding Statistical Errors in VA

\*\* We covered these chapters on Wednesday

# Chapters 5-11: Cautions & Recommendations

6. Shifting from School to Teacher VA
7. Marshaling the Evidence about VA
8. Evaluating VA Measures and Avoiding Double Standards
9. Using Value-Added to Improve Teaching and Learning
10. Creating Value-Added that Match Their Uses
11. Addressing Key Trade-Offs, Misconceptions, and Questions

\*\* We will cover these today

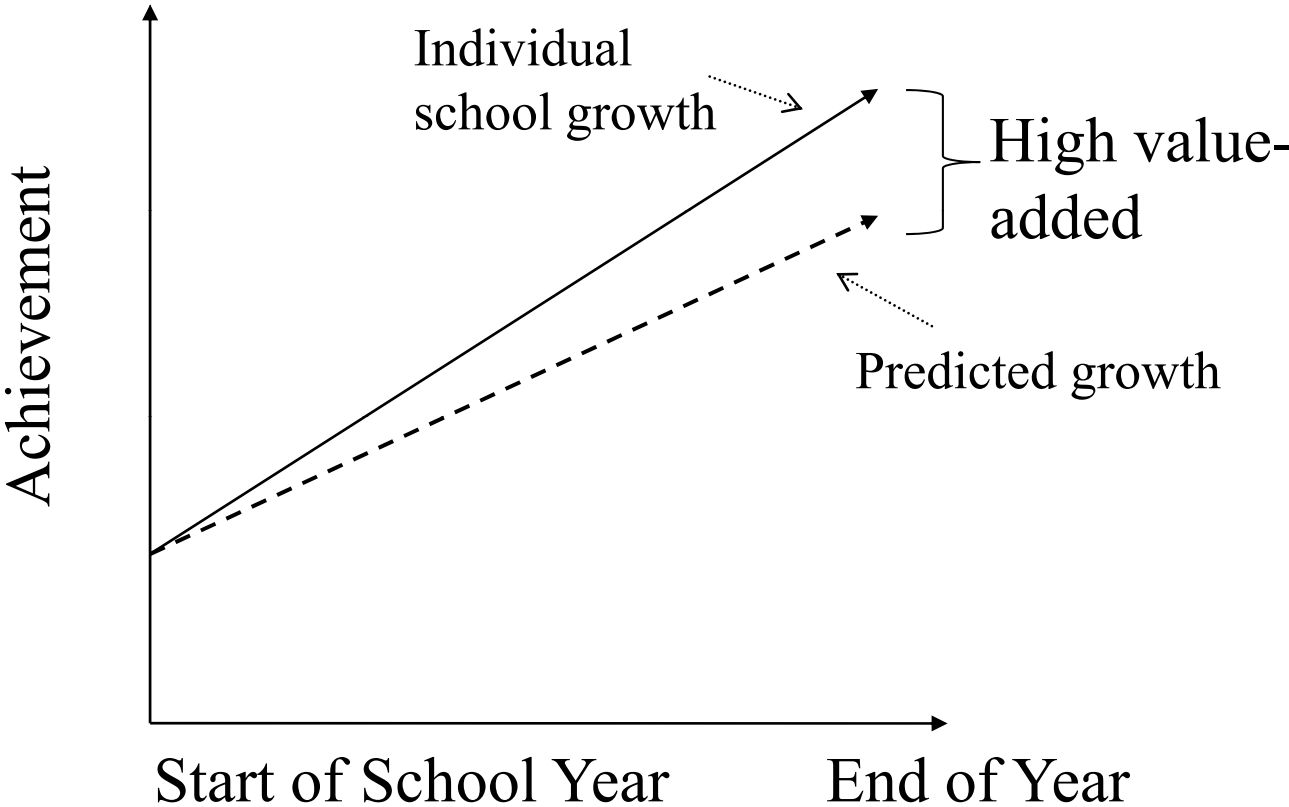


Chapter 6:  
Shifting from School to Teacher  
Value-Added

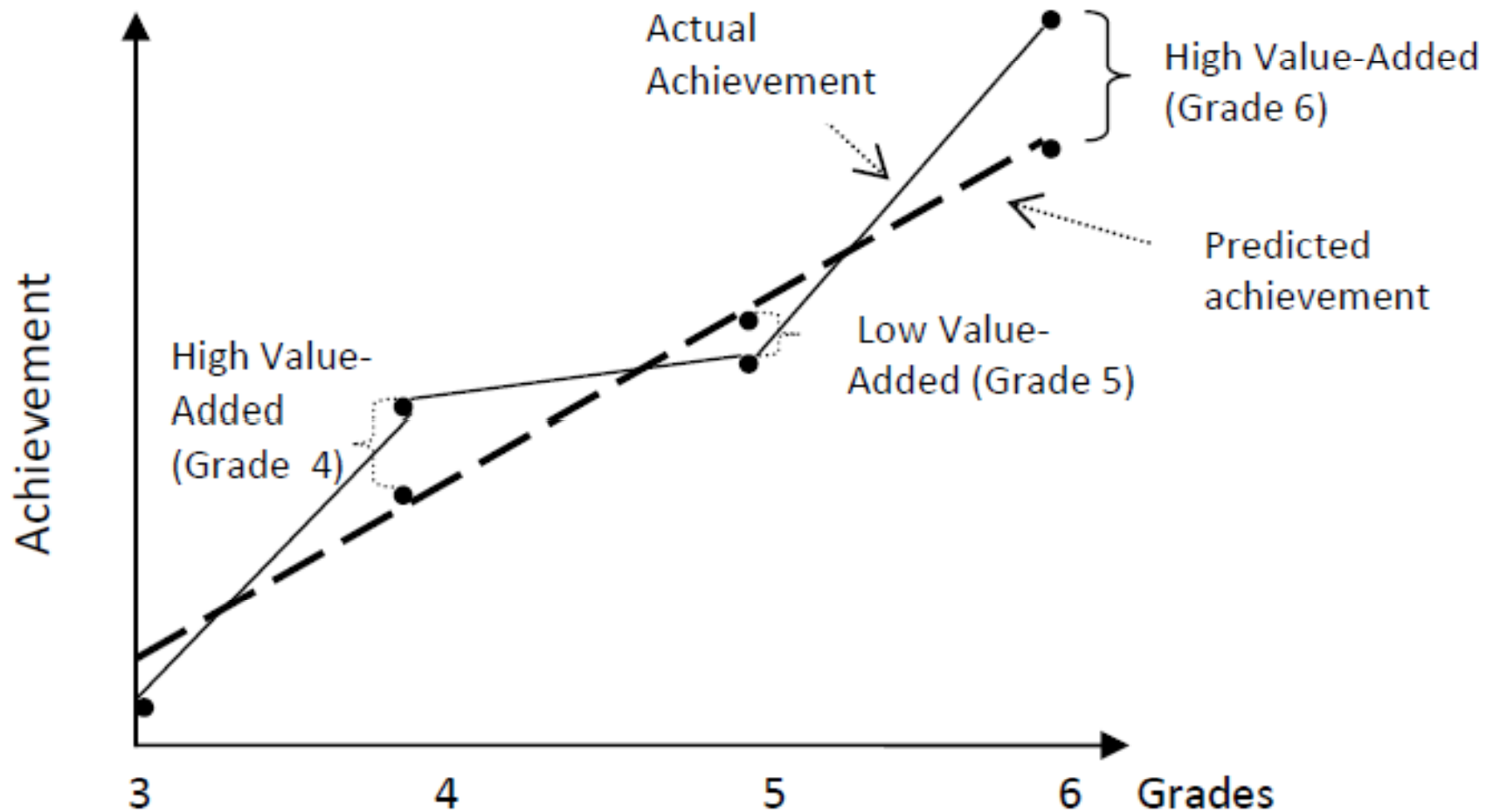
# More on Failure #1

- Problems with school value-added:
  - Does not focus in individual teachers
  - Lack of information for individual teachers (for personal accountability)
- *Apparently wide variation in teacher value-*
- How could we shift this from schools to teachers?

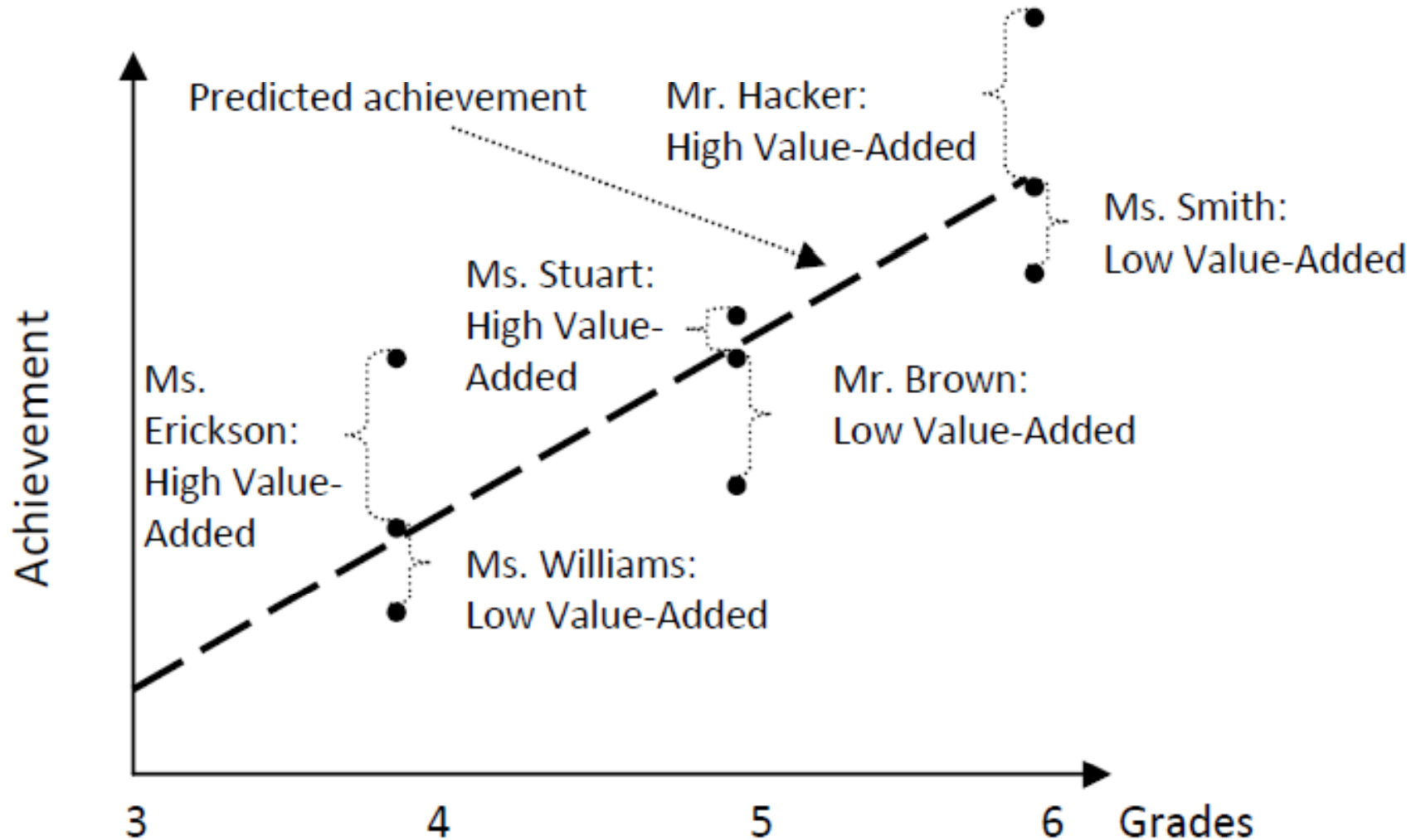
# Recall Basic VA Picture



# From Schools to Grades-within-Schools

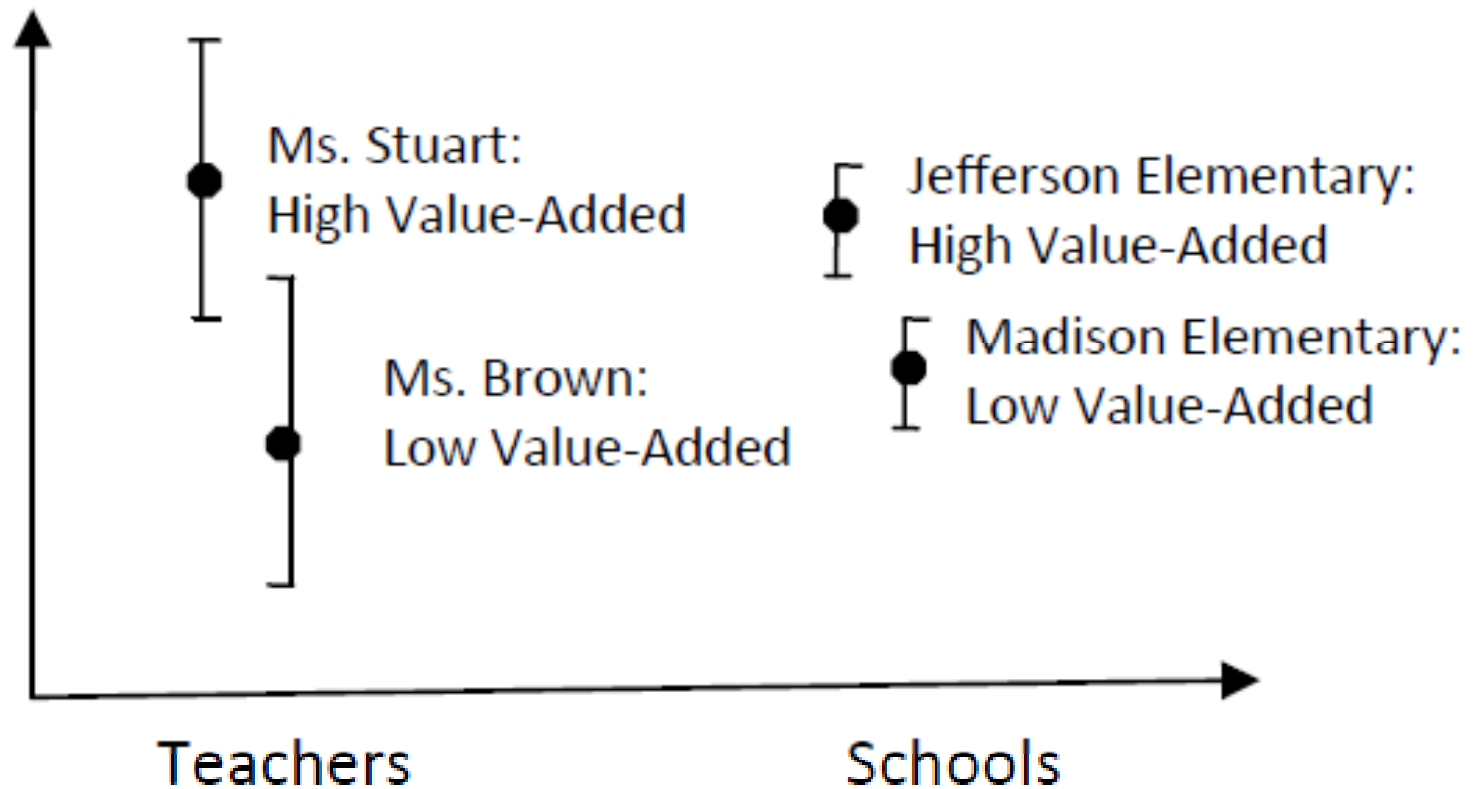


# From Grades to Teachers-within-Grades



# Problem: Few Students $\Rightarrow$ Random Error

Value-Added






# Different Grades and Subjects, Different Problems

	Advantages	Disadvantages
Elementary	Easier to attribute gains to teacher	Few students per teacher
Middle	Many students per teacher	Attribution to teacher more difficult (except math)
High	Many students per teacher	Less frequent testing and worse alignment to curriculum (except with end-of-course tests)

# A Middle Ground: Team VA

- School VA measures are more precise, but teacher VA is arguably what we really want
- What to do?
- One possibility: split the difference and evaluate teams of teachers within schools
  - Gets us closer to what we want and with greater precision
  - And arguably less bias because fewer unmeasurable factors introduce systematic error



# Chapter 7: Marshaling the Evidence about Value-Added

# Research on Teacher VA: The Good News

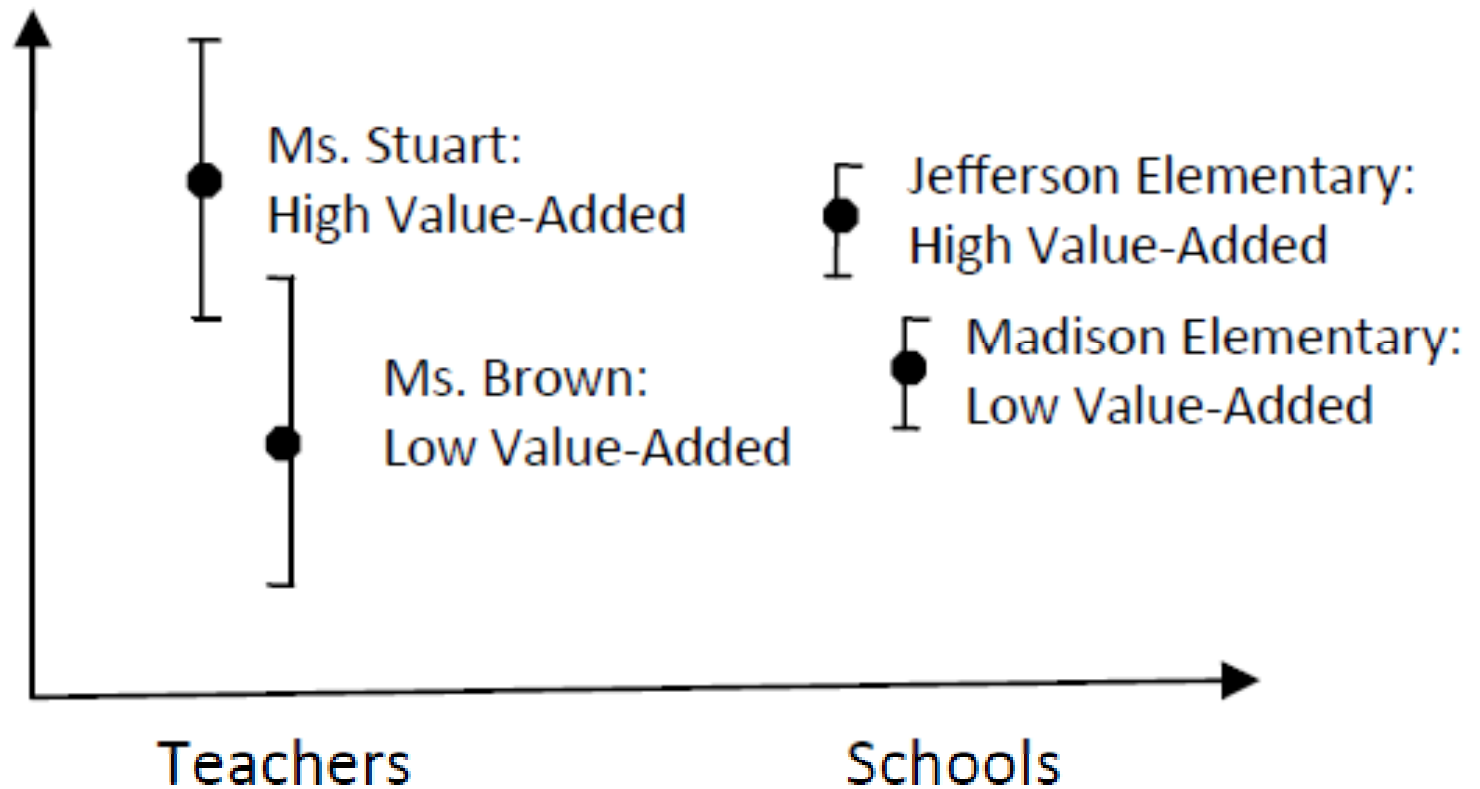
- Differences between lowest and highest value-added teachers seem large
  - Differences are exaggerated though
- VA measures have been partly validated by a random assignment experiment (Kane and Staiger in LA)
- VA measures of teacher effectiveness are positively correlated with principals' subjective assessments of teachers

# The Bad News

- Research focuses on what's typically true, but this means making a lot of specific mistakes
- VA no better than tests—garbage in, garbage out
  - Much effort right now toward improving the quality of student assessments
- VA measures vary across tests (same subject)
- Sensitive to specific statistical assumptions
- Are imprecise
  - Hard to say that one teacher is clearly better than another based on VAM-A
  - As a result, teacher measures are unstable

# Problem: Few Students $\Rightarrow$ Random Error

Value-Added




# Instability in Other Professions

- Brookings report says that stability of teacher value-added measures stacks up well to other professions, especially baseball
- But not clear how to interpret this: stability of baseball pitchers much different than baseball hitters
- Also, even in baseball, statistics are used in combination with a lot of other information to make decisions

# The Limited Applicability of VAM

- One of main limitations of VAM is that, in most states, it can only be applied easily in grades 4-8, math and reading
- Excludes:
  - Teachers in other subjects, coaches, specialists
  - Teachers in grades K-3 and 9-12
  - New teachers



Chapter 8:  
Evaluating Value-Added and  
Avoiding Double Standards

# Interpreting Evidence: Researcher Perspective


- Researchers are very conservative when it comes to drawing conclusions
  - We do not say that a measure (e.g., performance) is a good measure unless it is highly valid and reliable
  - Focus on statistical validity and reliability
  - See AERA/APA/NCME standards
- Critics of value-added have adopted a researcher stance (Hill, 2009)
  - “value-added<sub>ran</sub> kings [should] converge with results from other ratings of quality, such as classroom observations, parent surveys, ...”
  - “different student-level assessments within the same domain [should] yield same results”

## Interpreting Evidence: Policymaker Perspective

- Basic principle of policy analysis is to compare the viable options and choose the best one(s)
- Larger goal is to improve teacher effectiveness—is teacher VA among the best strategies?
- All measures have their advantages and disadvantages

# The Double Standard

- Critics of VA often don't apply same standard to credentials that they do to value-added
  - Example: Do credentials “converge with results from other ratings of quality, such as classroom observations, parent surveys, ...”?
  - Answer: No way.
- No performance measure could possibly meet the AERA/APA/NCME standards



Chapter 9:  
Using Value-Added to Improve  
Teaching and Learning

## Guiding Principle #2


- The stakes attached to a measure should be proportional to the quality of the measure

# Recommendations for Using VA M

- #1: Use value-added to measure school performance and hold schools accountable
- #2: Experiment with and carefully evaluate policies that use value-added to measure the performance of individual teachers
- #3: In creating performance measures, combine value-added with other measures more closely related to actual practice
- #4: Experiment with and carefully evaluate policies that use value-added to measure the performance of teacher teams

## Recommendations: Part II

- #5a: Consider extending value-added to other grades, subjects, and student outcomes . . .
- #5b: . . . But don't let the tail wag the dog.
- #6: Avoid the “air bag” problem. Don't drive value-added measures “too fast.”



Chapter 10:  
Creating Value-Added that Match  
Their Uses

# Recommendations for Creating and Reporting VA Measures: Part I

- #1: Use student tests that reflect rich content and are standardized, scaled and criterion-referenced
- #2: Create data systems that link student outcomes over time and to teachers and schools
- #3: Include all students, including special education, English Language Learners, and students with some missing data
- #4: Make adjustments to align the timing of the test with the timing of schooling activities

# Recommendations for Creating and Reporting VA Measures: Part II

- #5: Average value-added measures over  $\geq 2$  years
- #6: Create value-added measures based on comparisons among teachers and schools that facilitate cooperation and collaboration
- #7: Create value-added measures that compare teachers with grades and subjects
- #8: Account for factors that are outside the control of those being evaluated
- #9: Adjust for sampling error
- #10: Report confidence intervals

## An Additional Recommendation

- Use value-added to evaluate school, district, and state programs and practices
- The evidence on teacher credentials is a good example
- The value-added approach solves the same problem in program evaluation as it does in educator performance
  - Avoid systematic errors

# Is Missouri following these recommendations?

- Short answer: Yes
- Starting with district and school VA
- VA measures will be “shrunk”
- Used in a low-stakes way, as information
- >2 cohorts of students
- Some decisions still to be made

# Misconceptions: Part I

- (1) We cannot evaluate educators based on value-added because:
  - (a) Different teachers have different students
  - (b) Value-added measures have flaws
  - (c) Student tests are inadequate
  - (d) Teaching is complicated
  - (e) Student needs are diverse
- (2) Value-added is fair for teachers but not for students

## Misconceptions: Part II

- (3) Value-added measures are not useful because they are summative rather than formative
- (4) Because they involve comparing teachers to one another, and there are no absolute value-added standards, value-added measures are not useful for accountability
- (5) Because we know so little about the effects of value-added, we cannot risk our kids' futures by experimenting with it

# Misconceptions: Part III

- (6) Value-added is too complicated for educators to understand
- (7) Value-added simply represents another step in the process of “industrializing” education, making it more traditional and less progressive
- (8) Value-added is a magic bullet that by itself will transform education

# Moving Toward Teacher Effectiveness

- We can do better than the credentialing and checklist evaluation systems
- In deciding how to use VA, we should:
  - (1) Ask ourselves: Is this system going to give high ratings to the types of teachers and schools I would want my children to attend?
  - (2) Compare VA to the alternatives
- We need a comprehensive system of teacher effectiveness and performance measures in some form represent one important element



Questions?