Center for
**Educator Compensation Reform**

# Value-Added Measures of Educational Performance: Pitfalls and Promises

Robert H. Meyer, Director
Value-Added Research Center
Wisconsin Center for Education Research
University of Wisconsin-Madison

# Why Value-Added?
# Why not Average Attainment or the Proficiency Rate?
# Why not a Combination of the Two?

- Three **criteria** for evaluating models and indicators of school productivity. Meyer (1996, 1997)

# Three Criteria

- **Criterion validity/alignment**: Are VA school productivity indicators measured in terms of student outcomes valued by students and society?
- **Statistical**: Are VA indictors accurate in the sense of measuring true school productivity?
  - Unbiased? Precise (low error)? Minimum mean-squared error?
- **Behavioral**: Are the VA indicators non-corruptible?
  - Do VA indicators provide proper incentives to make decisions to maximize growth of student achievement.
  - Are all students included and are all students weighted equally? (the ”one person-one vote rule” applied to educational accountability).
  - Are control variables fixed or subject to manipulation by schools? (Example: gender and race/ethnicity are largely fixed student attributes, whereas participation in special education is both a student attribute and a factor that is partly determined by educational staff.)

# Why not Average Attainment or the Proficiency Rate? – Statistical

- Biased as measures of school productivity, even if they are derived from highly valid assessments.
- Attainment indicators are biased because they:
  - Reflect prior achievement and family and student factors associated with achievement growth
  - Reflect out-of-date productivity effects from prior grades and years (back to pre-school and early grades)
  - Are contaminated due to student mobility (and the bias differs across schools)
  - Fail to localize school productivity to a specific grade level, but rather capture (at best) productivity effects from pre-school and onward.

# Why not Average Attainment or the Proficiency Rate? – Behavioral

- Provide institutions with the perverse incentive to "cream," that is, to raise measured performance by educating only those students that tend to have high test scores.
- Creaming mechanisms:
  - Selective admissions
  - Create an environment (not necessarily intentionally) that is unsupportive to potential dropouts, academically disadvantaged students, and special education students
  - Aggressively retain students
  - Migration of high-quality teachers and principals to schools with academically advantaged students

# NAEP Mathematics Examination Data

**Average Test Scores by Year (a)**

| Grade | 1973 | 1978 | 1982 | 1986 |
|---|---|---|---|---|
| 3rd | 219.1 | 218.6 | 219.0 | 221.7 |
| 7th | 266.0 | 264.1 | 268.6 | 269.0 |
| 11th | 304.4 | 300.4 | 298.5 | 302.0 |

**Average Test Score Gain From Year to Year for Each Cohort (b)**

| Grade | 73 to 78 | 78 to 82 | 82 to 86 |
|---|---|---|---|
| 3rd to 7th | 45.0 | 50.0 | 50.0 |
| 7th to 11th | 34.4 | 34.4 | 33.4 |

**Source**: Dossey et al. (1988).

**Figure 1. A Graph of Student Achievement Data for Two Schools**

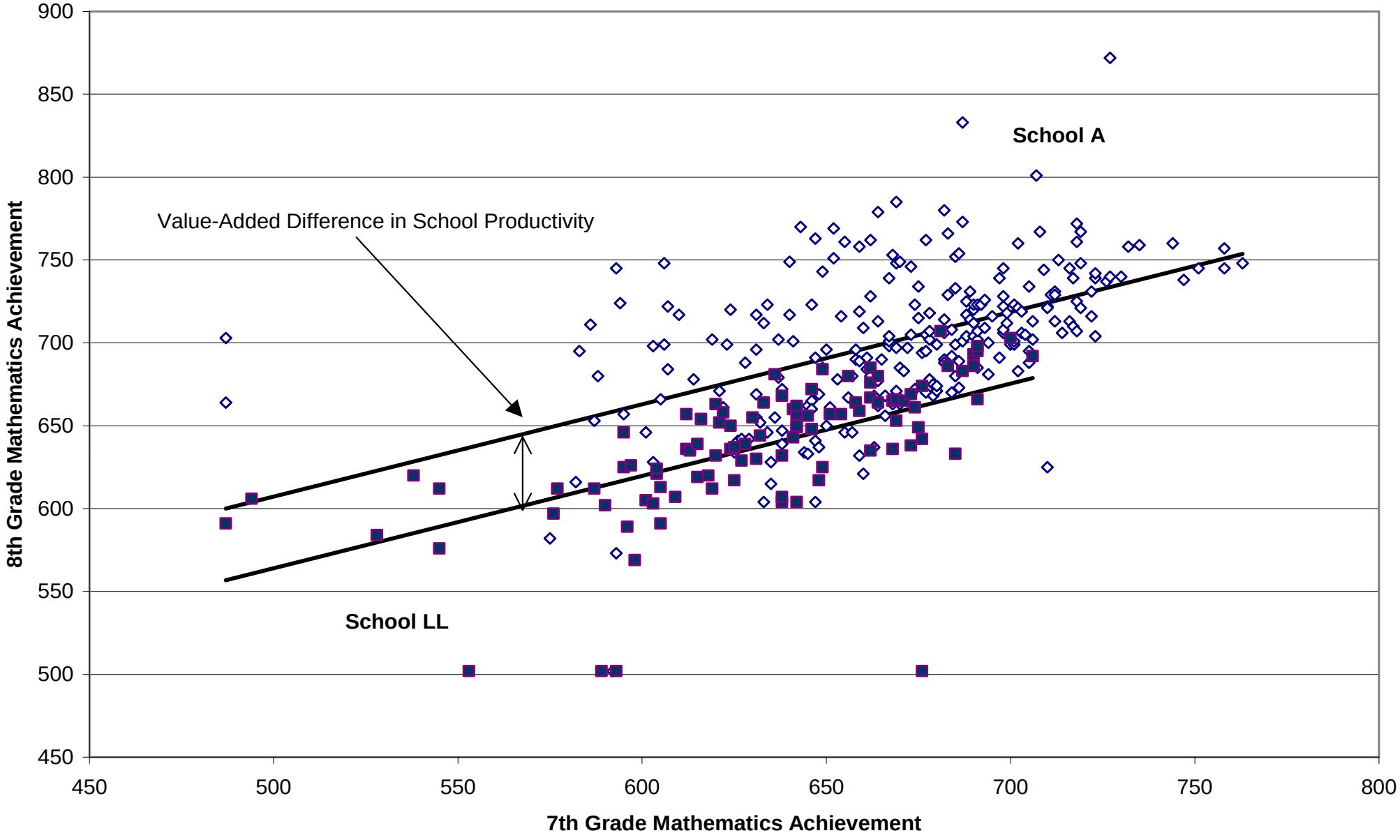
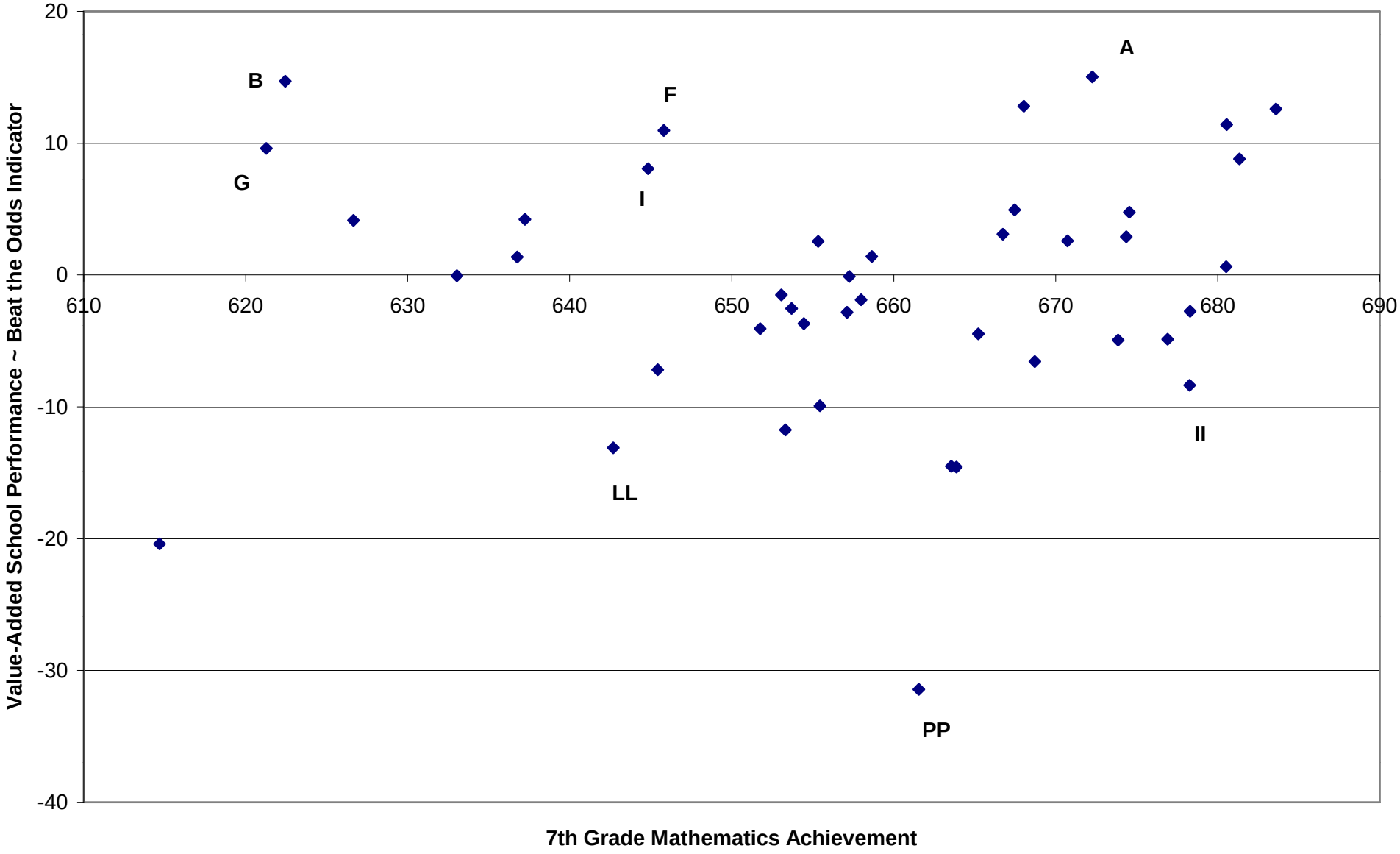
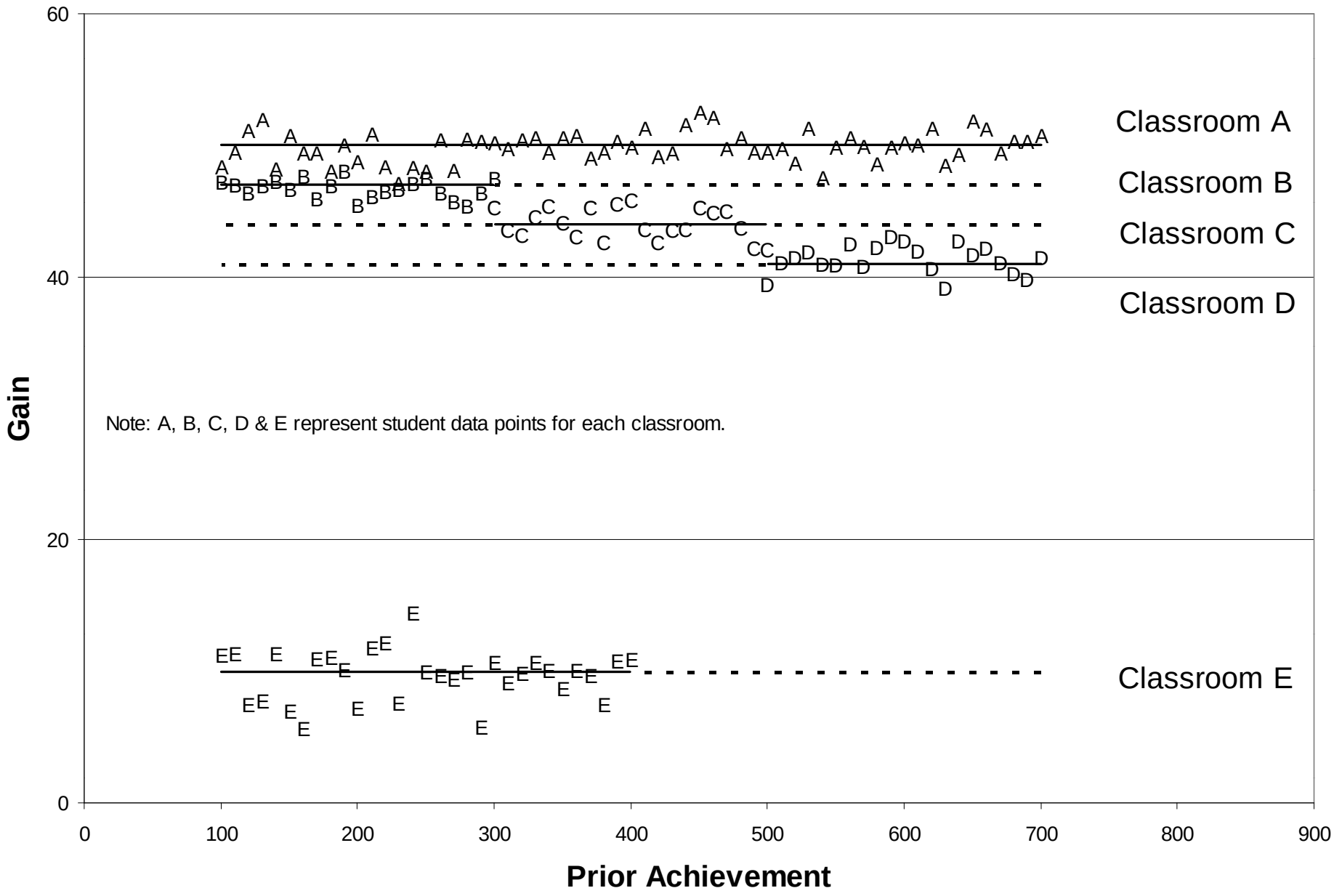


**Figure 3. School Performance by Average Prior Achievement**



# Figure 1. The Value-Added Productivity of Classrooms With Different Student Populations



# A Value-Added Model of 4th Grade Student Achievement: Statistical Specification

$$\text{Posttest}_{4i} = \lambda \text{Pretest}_{3i} + \beta \text{Student Characteristics}_i$$
$$+ \delta \text{School Factors}_{classroom,school}$$
$$+ \text{VA}_{classroom,school} + \text{Unknown Student Factors}_{4i}$$

# Construction of VA Indicator: It is the part of growth left over after controlling for other factors

$$\text{VA}_{classroom,school} = \text{Classroom Average}\{\text{Posttest}_{4i}$$
$$-\lambda\text{Pretest}_{3i} - \beta\text{Student Characteristics}_i$$
$$-\delta\text{School Factors}_{classroom,school})$$
$$+\text{Classroom Average}(\text{Unknown Student Factors}_{4i})$$

**Q**. What does a classroom/teacher VA indicator measure?

**A.** The combined productivity of teacher, classroom, principal, and school inputs. Thus, it is a proxy measure of teacher productivity.

# Issues in the Development of a Value-Added Indicator System

**How complex should a value-added model be?**

- Possible rule: "Simpler is better, unless it is wrong."
- Wrinkle: Greater complexity may come at the expense of statistical precision (reliability). Hence, the decision to adopt a more complex model may necessitate greater aggregation (over grades and/or years).

(1) **Test Date**

Do you test students only near the end (early May) or beginning (late September) of the school year? If so, do you want to take account of the fact that the annual growth period (say, March to March) cuts across two school years and, typically, two different teachers or sets of teachers.

## (2) **Demographic controls**

Control for differences across schools in student demographic characteristics?

- Income status (free lunch)
- race/ethnicity
- gender
- special education
- English language learner, bilingual

(3) **Retention** Include retained-in-grade and promoted students in the estimation of school effects? (Almost certainly yes.)

(4) **Student mobility** Include students who changed schools over the summer (that is, within the annual testing interval if tests are not administered near the end or beginning of the school year)? (Probably yes.)

## (5) **School-year mobility**

Include students who changed schools during the school year *and* take account of within-school year mobility by defining school enrollments in the model as the fraction of the school year enrolled in a given school (dose model)?

## (6) **Classroom/teacher indicators**

What does a classroom teacher indicator represent? Answer: The combined productivity of teacher, classroom, principal, and school inputs. Thus, it is a proxy measure of teacher productivity.

## (7) **Measurement error in test scores**

If prior achievement is a control variable in the value-added model, do you control for measurement error in this variable? (Almost certainly yes, because failure to control for measurement error yields biased parameter estimates.)

(8) **Aggregation over units: schools, schools by grade, teacher teams, individual classroom /teacher/ school?**
Statistical precision is highest at the highest level of aggregation since precision increases with the number of students.

- Where should incentives be directed: at individuals or teams?

(9) **Aggregation over time.** “Smooth” data over time to improve precision?

(10) **Multiple components.** Separately estimate the productivity of regular school (and teachers), summer school, after school, NCLB Supplemental Education Services (SES)?

(11) **Special education detail.** Control for many different types of special education status (type and severity of handicap)?

(12) **Multi-year data.** Exploit multiple years of longitudinal student data to implicitly control for heterogeneity in student achievement growth profiles?

# Alignment of Value-Added Measures with Subjective Ratings of School Productivity

| **Subjective Measure of Productivity** | **Expected Correlation with Value-Added Indicator** |
|---|---|
| I. Productivity measured in a manner intended to parallel productivity as measured with respect to student achievement | High |
| II. Productivity measured to capture dimensions not captured by student achievement | Medium |

# Contact Information

**Robert H. Meyer**
**Value-Added Analytics: *VA2***

**http://varc.wceruw.org/**

**RHMeyer@wisc.edu**