Operation Public Education

Value-Added Assessment
Achievement -- a score on a vertical scale at a single moment in time (absolute or raw score, status, proficiency) -- is best predicted by family income.
Family Income and SAT Scores

**2005 College-Bound Seniors: Total Group Profile Report** published by CollegeBoard SAT
New technologies and data sets make new research possible

- James Coleman (1966) and Christopher Jencks (1972) concluded that family background was more important than schooling in explaining achievement.
- They did not have the technology to trace individual students over time (not cohorts).
- Nor did they have data sets that link the teacher for every subject and grade to individual student’s record.
Each child serves as his own statistical control

The Environmental Variables Remain the Same
- Family Income
- Ethnicity
- Gender
- Neighborhood

Grade 3
Grade 4
Does the Percentage of Students Receiving Free and Reduced-Price Lunches Affect System Gains?

Cumulative Gain of a Large East Coast County’s School Systems Compared with the Percentage of Students Receiving Free and Reduced-Priced Lunches

Reading

Income has no effect on value-added
Growth (the progress students make over the course of the school year) is best predicted by the quality of instruction.
Identifying AYP’s shortcomings

<table>
<thead>
<tr>
<th>Achievement</th>
<th>Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Achievement</td>
<td>High Growth</td>
</tr>
<tr>
<td>Low Growth</td>
<td></td>
</tr>
<tr>
<td>Low Achievement</td>
<td>Low Growth</td>
</tr>
<tr>
<td>Low Achievement</td>
<td>Low Growth</td>
</tr>
<tr>
<td></td>
<td>High Growth</td>
</tr>
</tbody>
</table>
EVAAS was developed for Tennessee by William Sanders since 1992, tracks each of the state’s 885,000 students. It is the largest data base ever assembled and is mandatory in Pennsylvania and Ohio as well as in over 300 districts and consortia across the U.S. Other value-added models exist, but Sanders’ is currently the only one used on a statewide basis.
Philosophy Behind Value-Added Assessment

- Schools can and should add value for each student from September to June.
- This is true whether the student comes in above grade, at grade or below grade.
- Students are entitled to grow at least at a rate they have demonstrated in the past.
Don’t confuse value-added assessment with simple growth or gain

- “Growth” or “gain” is calculated by subtracting last year’s score from this year’s score on vertically-scaled tests.
- “Simple” and “transparent” approaches such as this produce specious and erratic results.
- Value-added assessment is more powerful than a simple growth or gain score.
EVAAS

- Analyzes the difference between “projected” and “actual” scores
- Estimates how much of the difference between these is attributable to the student and how much is attributable to the teacher
- Arrives at a score for each teacher measuring the effectiveness of their instruction
Value-added measures the difference between actual and projected.
Value-added measures the difference between actual and projected.
What makes value-added fair?

1. For children
   Value-added is fair to students because it bases their projected score only on their prior academic record. That ensures that all children are expected to make progress each year from wherever they start.
What makes value-added fair?

1. For educators
   It is fair to administrators and teachers because prior academic achievement data already incorporate the student background characteristics that bias absolute test scores.
Defining a successful school

- Each year the performance of the students exceeds what is expected of them, given their academic background.
- Over time all students are able to achieve high standards (NCLB).
Value-Added: The Basics

- Value-added is not a test.
- It is a way of looking at the results that come from tests.
- Value-added lets us determine whether the students in a class, school or district are making enough academic growth each year.
Value-added provides powerful diagnostic data

- Identify and improve the focus and impact of instruction
- End the isolation of teachers
- Build learning communities
- Improve data-driven decision making
- Differentiate instruction
- Create student growth trajectories to targets and develop intervention strategies
- Measure the success of schools through growth, not simply achievement
Value-added assessment is only a thermometer; if we don’t analyze the information and use it, nothing happens.
Value-added as a diagnostic is no panacea
NAEP scores: Tennessee and the nation

Mathematics - Grade 4

NAEP Scale Score


Tennessee National Average

- Tennessee
- National Average
The limits of value-added in Tennessee

- May be used in individual teacher evaluations, but may not exceed 8%
- Lack of professional development to accompany statewide rollout
There are some very good schools and some very poor schools, but it is impossible to determine where a school falls just by knowing its location or the make-up of its student body.

Math: 1996-97

- 166 schools with % of National Norm above 110
- 250 schools with % of National Norm 100-110
- 330 schools with % of National Norm 90-100
- 234 schools with % of National Norm 80-90
- 108 schools with % of National Norm 70-80
- 121 schools with % of National Norm below 70
Why collect classroom level data?

- The variation in the quality of instruction is much greater within schools than between schools
- Struggling students are not randomly distributed in classrooms – they are found disproportionately in classrooms where they receive poor instruction
- Deal with underlying causes not symptoms
Value-Added Findings From Tennessee

The Teacher Effect
Probability that a bottom-quartile 4th grade student will pass the high-stakes graduation exam in 9th grade

Poor teacher sequence: <15%
Average teacher sequence: 38%
Good teacher sequence: 60%
Teacher Effectiveness

Teacher Experience

First 10-12 years

Second 10-12 years

After 20-24 years

Typical Salary Schedule
For use in teacher evaluations

- Part of a balanced system: observation (inputs) and empirical (outputs)
- Multiple measures: planning and preparation, classroom environment, instruction, professional responsibilities
- Appropriate safeguards, e.g. peer assistance and review panels
- Expanded professional development (mentors, coaches, additional days)
Careful use of ranking data

- Teachers are ranked by their VA scores
- Two lines are drawn to separate teachers at both ends of the distribution to identify teachers who significantly different from the average teacher
- Three categories of instruction emerge: highly effective, effective, ineffective
- No differentiation within categories
Use of Value-Added Assessment in Comprehensive Reform

Highly effective:
Above

Effective:
No Detectable Difference (NDD)

Ineffective:
Below

One year’s worth of growth (using 3-year running averages)
For additional information on our package of reforms, please contact:

cgpinfo@pobox.upenn.edu
or (215) 746-6478

Or see our website at http://operationpubliced.org
Minority status has no effect on value-added
The Importance of Teacher Sequence