How Should Educators Interpret Value-Added Scores?

Stephen W. Raudenbush
Marshall Jean

University of Chicago
Outline

• Background
• The Problem: Decisions under uncertainty
• Comparing Individual Teachers
• Identifying Groups of Teachers
• Guidelines
Background

What is a Value Added Score?

Average learning gain
or “Adjusted” average gain

Who Might Use Value Added?

Teachers
Administrators
Parents
Researchers

The Problem:

We measure value added with error
Sources of Error

**Bias** (see McCaffrey’s entry)

Bias is minimize if teachers being compared teach similar students

**Imprecision**

number and consistency of test items
number of kids in the teacher’s classroom
“Where do I stand?”

Figure 1
Reliability

How precisely do we measure each teacher’s VA?
How much do teachers vary?

Reliability is

\[
\frac{\text{Variation of "true VA"}}{\text{Variation of "true VA"} + \text{Variation of the measurement errors}}
\]
Suppose we had more precision
Suppose teachers varied more
Suppose teachers varied more
Identifying Groups of Teachers

EG

- Superintendent wants to commend top 10 percent
- Or identify lowest 25% for extra help

Two kinds of error

- False identification
- False non-identification
Example

• We want to identify lowest 25%
• Correlation between two years of VA=.40
• We are willing to tolerate 50% false identification (?!)
• Then we pick lowest 16%
## Simulated Results

<table>
<thead>
<tr>
<th></th>
<th>(1) Truly below 25&lt;sup&gt;th&lt;/sup&gt; percentile</th>
<th>(2) Truly above 25&lt;sup&gt;th&lt;/sup&gt; percentile</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Estimated to be below 25h percentile</td>
<td>80 Correct</td>
<td>80 Falsely Identified</td>
<td>160</td>
</tr>
<tr>
<td>(2) Estimated to be above the 25&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>170 Falsely not Identified</td>
<td>670 Correct</td>
<td>850</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>750</td>
<td>1000</td>
</tr>
</tbody>
</table>
Can We Use These Numbers

No!
- Errors of Classification are Shocking
- High stakes use is arbitrary

Yes!
- VA more informative than seniority, degrees
- Kids of the 160 “low” teachers score 1 sd below average
- Very few of the 160 are in the top 25%
Conclusions

1. In comparing teachers,
   • Never rely on a “point estimate”
   • Use a confidence interval instead

2. In identifying sub-groups
   • Analyze risk of false identification
   • Analyze risk of false non-identification

3. Weigh tradeoffs between teacher rights and children’s welfare

4. Get more information!