

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

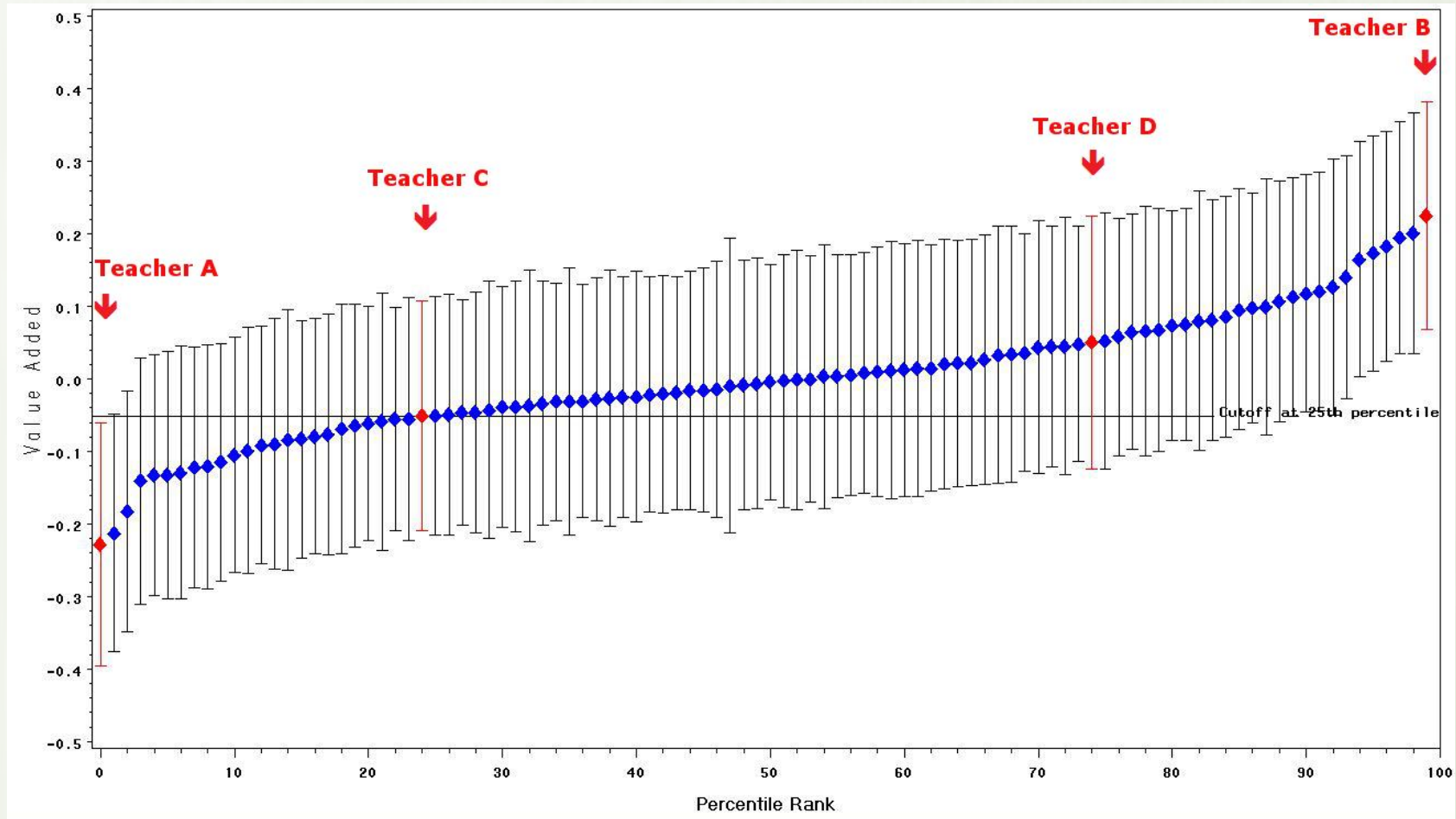
Imprecision

number and consistency of test items

number of kids in the teacher's classroom

“Where do I stand?”

Figure 1



Reliability

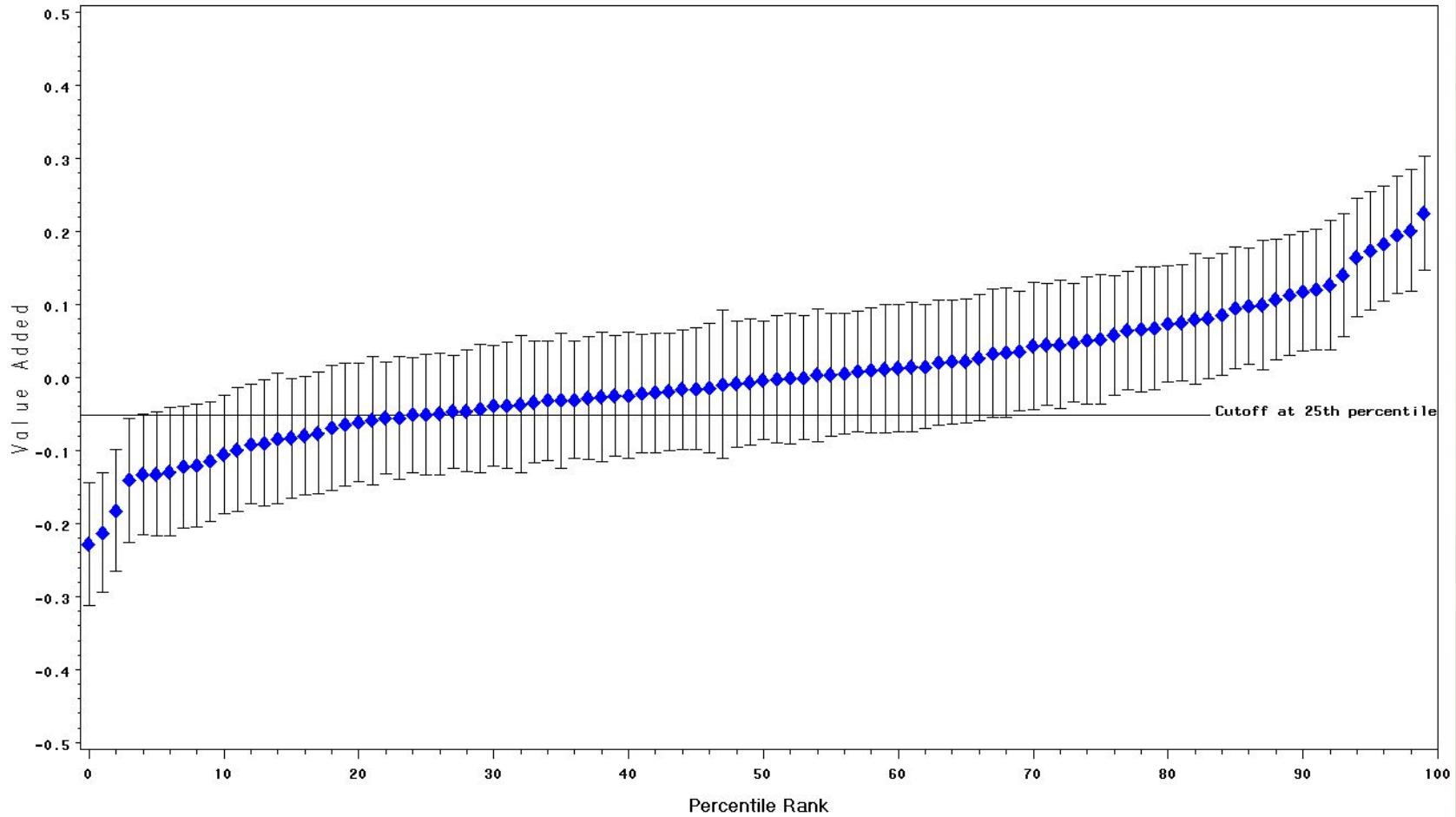
How precisely do we measure each teacher's VA?

How much do teachers vary?

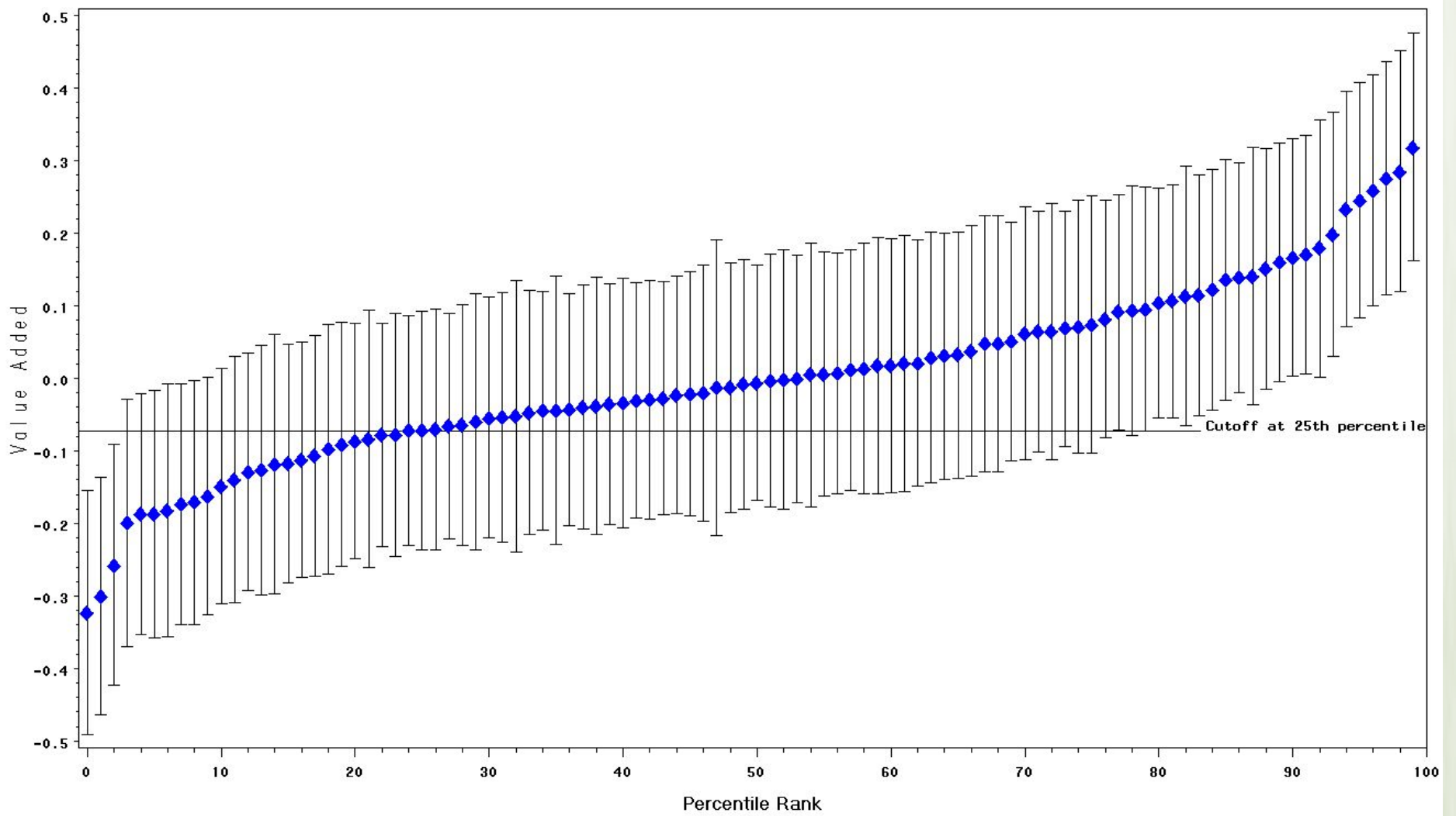
Reliability is

$$\frac{\text{Variation of "true VA"}}{\text{Variation of "true VA" + Variation of the measurement errors}}$$

Suppose we had more precision



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

	(1) Truly below 25 th percentile	(2) Truly above 25 th percentile	Total
(1) Estimated to be below 25 ^h percentile	80 <i>Correct</i>	80 <i>Falsely Identified</i>	160
(2) Estimated to be above the 25 th percentile	170 <i>Falsely not Identified</i>	670 <i>Correct</i>	850
Total	250	750	1000

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!