Using Value-added to Compare Teachers Who work in Different Schools

Stephen W. Raudenbush
University of Chicago
September 11, 2013

A webinar sponsored by the Carnegie Foundation for the Advancement of Teaching
Outline

• Challenges arise when comparing teachers who work in different schools
  – Variation in school effectiveness
  – Variation in contextual conditions
  – Non-comparability of students

• Checking sensitivity of value-added scores

• How to make more meaningful comparisons
What is Bias in VA?

An unbiased measure does not favor teachers

... who are assigned fast-growing students

... who work in resource-rich schools
Some Reasons to Worry About Bias

Suppose we rank teachers on value-added who

…work in the same school

…work in different schools

Will the rankings be similar?

Special concerns about comparing teachers who work in different schools

1. Schools vary in effectiveness
   • It’s an old debate among researchers!
   • New evidence from longitudinal studies
   • New evidence from randomized lotteries

But are schools more than the sum of their teachers?
   • Randomized studies of curricula
   • Randomized studies of school-wide reform
   • Studies of teachers moving schools
In sum….

Value-added studies confound teacher and school effectiveness

Good reason, then, to suspect that school effectiveness biases comparisons between teachers using VA
2. School context influences instruction and learning

- Highly motivated parents, peers
- Teachers tailor cognitive pace and conceptual level to prior achievement
- Neighborhood safety supports attention

Current value-added technology cannot separate impact of these from the impact of teacher skill
3. Statistical adjustment

Consider the problem of comparing two teachers

Rationale:
- Students are not assigned at random to classrooms
- Idea is to compare students in one classroom to similar students in other classrooms

Can work well if
- You have a very good pre-test
- You can find comparable students in both classrooms!

Can work badly if
- Many students in one classroom have no comparable peers in the other classroom
- Statisticians call this lack of common support
- We may be comparing teachers who are doing very different kinds of work
When is a Failure of Common Support Most Likely?

In Elementary Schools

- Most elementary children go to school near home
- Schools will often serve neighborhoods that are segregated by income or ethnicity
- Then teachers in different schools may serve very different children

In Secondary Schools

- Schools may serve different neighborhoods
- But large high schools often take in heterogeneous kids
- Yet classrooms within these large secondary schools may be “tracked.”
- Hence teachers in the same school may serve very different kids
Empirical Evidence:

Several studies exploit randomized experiments to test the bias of value added.

Results are encouraging.

Yet the random assignment occurred within schools.

One large-scale study looked at school mean achievement after a teacher with high value-added left the school.

“High value added” predicted earnings!
How Can We Proceed?

We can rank teachers on value-added who
...work in the same school
...work in different schools

Ask: Are the rankings similar?
In addition...

If we are willing to assume that
    .... We have good pre-test data
    .....and comparisons would be fair if common support is adequate,

Then we can compute an “upper bound” on the bias that arises from a failure of common support.
When will teachers look non-comparable?

When the mean pre-test varies significantly between classrooms

and...

There is a strong “contextual effect,*”

Then,

The risk of bias is high

*A “contextual effect” is in fact a statistical association between the classroom mean pretest and the outcome even after controlling for the individual-level pretest
What if the risk of bias is high?

One might then divide schools (or classrooms within a school) into subsets that serve similar students.

This changes the question that value-added scores answer, but

– It reduces the risk of bias

– Confines us to a question about which our data have information!
Overall Summary

Comparisons among teachers makes more sense when
– Teachers work in similar school environments
– And are teaching reasonably similar students

To some extent, we can check our data to see if these conditions hold

If they do not, we should take steps to insure that comparisons are meaningful – by finding subsets of teachers who are teaching similar students in similar conditions