

# Does Value-Added Work Better in Elementary Than in Secondary Grades?

Douglas N. Harris  
Associate Professor of Economics  
University Endowed Chair in Public Education  
Tulane University

Carnegie Foundation webinar  
June 7, 2013

# Teacher Performance Measures and School Context

- No performance measure is likely to work equally well in all circumstances
  - By “work well” I mean have high validity and reliability and therefore few “misclassification errors”
- Here focus on how well value-added (VA) measures work across grade levels
  - Most research on VA is focused on elementary schools, but there are reasons to think their properties differ in secondary grades
  - Here, I present new evidence

# Questions

- What do we know about how well value-added measures work across grade levels?
- What more needs to be known on this issue?
- How, and under what circumstances, does this issue impact the decisions and actions that districts can make on teacher evaluation?

# Well Known Differences Between Elementary and Secondary Schools

- In most elementary schools . . .
  - Most students have just one teacher, which makes it easier to attribute student achievement to one person
  - But each teacher has fewer students
- In secondary schools . . .
  - Students have many teachers and teacher effectiveness may spillover across subjects—for example, the social studies teacher might influence reading skills in ways that affect reading and writing scores (validity perhaps worse)
  - But more students per teachers (reliability perhaps better)

# Another Big Difference: Tracking

- In elementary schools, students are often evenly distributed across classrooms
- In secondary schools, students are tracked into courses by ability
- Tracking creates two main problems:
  1. Course content aligns differently to test across tracks (part of this is the well known ceiling effect problem)
  2. “Better students” sort into higher tracks
- Problem #1 biases VA in favor of lower-track teachers, while problem #2 biases in favor of upper-track teachers
- Not clear what the net effect is in terms of which teachers are advantaged, but we will examine this shortly

# Possible Solution to Tracking Problem

- Many administrative data systems include indicators for tracks, so could include the track indicators in the value-added model
- This is analogous to including student background information to account for measureable student differences
- Most value-added systems don't include tracks though there has been renewed attention to this

# What Happens When We Include the Track Indicators?

- The teacher value-added estimates change substantially
- We estimated teacher value-added for teachers with and without accounting for tracks
- A teacher with all lower-track courses who is really at the 50<sup>th</sup> percentile gets a VA score at the 30<sup>th</sup> percentile (on average)
  - This suggests that the net effect of excluding tracks is to bias VA in favor of upper-track teachers
- Only 25-50 percent of teachers are in the same performance quartile when comparing VA estimates made with and without course indicators

# What About Teacher Sorting?

- Teachers also sort by tracks, so including course indicators may “over-correct” and still leave a bias
- More experienced teachers in our sample are more likely to teach in the upper track; if these teachers have higher “real” value-added, then controlling for tracks will put them at a disadvantage
- Solution: Results reported earlier are based difference between value-added of lower- and upper-track teachers who teach in both tracks
- Above solution can also work for estimating VA for all teachers (not just those in multiple tracks) through a two-stage estimation process; discuss with your vendor



# What About End-of-Course (EOC) Exams?

- The tracking problem is in some ways less significant with EOCs because this improves alignment across all the tracks
- Some worry that we cannot calculate real “growth” with EOCs because the prior year score is not usually the same subject, but this is a problem with other tests as well
- Also, the prior score in value-added may largely capture general student skill and motivation, in which case the lack of a “pure” pre-test with EOCs is not a problem

# What About *Reliability* Across Grades?

- At first, it would seem that secondary teacher value-added will be more reliable because there are more students per teacher
- But this assumes that: (a) the variance in true teacher value-added is the same across levels; and (b) that there are no other factors, such as test measurement error, that also influence the confidence intervals
  - Reliability is the ratio of true variance to the variance of the estimate
- Secondary teacher value-added seems no more reliable, though this is likely to vary across states

# Predicting Future VA

- Several recent studies have focused on how well past performance measures predict future performance
- Evidence from North Carolina (with EOCs) suggests that prior VA is a worse predictor of future VA in the secondary level than it is at the elementary level
- This is true even after accounting for tracks
- Suggests that, overall, VA works better at the elementary level, so the evidence on VA is probably too optimistic

# What more needs to be known on this issue?

- Again, the vast majority of research on the validity of VA is at the elementary level, so we need to extend those types of studies to secondary grades
- It is not clear whether adding indicators is sufficient, partly because the course information in data systems do not always match the content
- Also, what about “non-standard” courses?

# How does the evidence and discussion here impact decisions?

- Policymakers should consider using different methods to estimate value-added at different grade levels
- They might also consider attaching different weights to value-added
- Some might argue that we cannot treat teachers at different grade levels unequally, but *equal* treatment is not the same as *equitable* treatment
- Also need to be leery of letting the value-added system drive how teachers are assigned to classes

# Summary

- The vast majority of research on value-added measures focuses on elementary schools; value-added measures for middle and high school teachers pose particular challenges.
- Middle and high schools often “track” students in ways that affect the validity of value-added.
- Student tracking in middle and high schools calls into question the validity of methods typically used to create value-added measures.

# Summary

- The validity of secondary-level value-added measures can be improved by directly accounting for tracks and specific courses, although this may not completely solve the problem.
- Middle and high school teachers have more students, and this factor increases reliability, but it is offset by other factors that reduce reliability at those grade levels.
- End-of-course exams, which are becoming more common in high school, have both advantages and disadvantages for estimating value-added.