

Is It Just a Bad Class?

Assessing the Potential of Using Value-Added Estimates (VAMs) of Teacher Job Performance for Making Tenure Decisions

Dan Goldhaber & Michael Hansen*

University of Washington Bothell

CALDER

November 21, 2008

The full report is available online at

www.crpe.org

*The research presented here is based primarily on confidential data from the North Carolina Education Research Center (NCERDC) at Duke University, directed by Clara Muschkin and supported by the Spencer Foundation. The authors wish to acknowledge the North Carolina Department of Public Instruction for its role in collecting this information. We gratefully acknowledge the Institute of Educational Studies at the Department of Education for providing financial support for this project. This paper has benefited from helpful comments from participants at the APPAM 2008 Fall Research Conference, the University of Virginia's 2008 Curry Education Research Lectureship Series, and the Economics Department Seminar Series at Western Washington University. We also wish to thank Carol Wallace for editorial assistance. The views expressed in this paper do not necessarily reflect those of the University of Washington or the study's sponsor. Responsibility for any and all errors rests solely with the authors.

Using VAMs to Assess Teacher Performance

- Teachers are the most important schooling factor explaining variation in student achievement, but ...
 - (Easily quantifiable) teacher characteristics used to determine teachers' employment eligibility and compensation don't strongly predict teacher effectiveness
 - Even when there are statistically significant differences, the differences between the best and worst teachers who hold a particular credential swamp the differences between those with and without the credential

Using VAMs for Policy Purposes

- Pay, tenure, and teacher “de-selection” reforms
 - Tennessee and Dallas using individual teacher as unit of analysis
 - Pay-for-performance in Florida, Texas, and Minnesota; TIF grantee districts
 - New York City vs. New York State on student test scores
 - De-selection/selective retention ideas associated with researchers (Gorden et al., 2006; Hanushek, forthcoming)
- Underlying tenure/de-selection is the notion that teacher quality is relatively stable characteristic

But... Significant *Potential* Problems with VAMs

- Logistical issues (for example, timing of tests)
- Perverse incentives/unintended consequences (such as reclassification of students, a too-narrow focus on tested items, or discouraging collaboration)
- Theoretical/practical issues measuring teacher contributions (cross-subject complements)
- Defining the constructed counterfactual (for instance, within or between school/district comparisons)
- **Measurement issues/stability of teacher performance**
 - Signal-to-noise ratio
 - Year-to-year changes in estimated performance
 - Sensitivity of performance ranking to changes in sample, subject, or teaching context

Background

- On teacher performance estimates:
 - VAMs can be biased by failure to correct for non-random distribution of teachers and students (Clotfelter et al., 2006)
 - VAM models on some specifications fail falsification tests (Rothstein, 2008)
 - Experimental VAM estimates look similar to non-experimental estimates with particular specifications (Kane and Staiger, 2008)
- On stability of performance estimates:
 - Year-to-year correlation of teacher effects fall in the 0.2 to 0.5 range; significant movement across quintile rankings between adjacent years (Aaronson et al., 2007; Koedel and Betts, 2006; Lockwood et al., 2008)

Thoughts on VAMs in Practice

- For policy purposes we probably don't care about precise estimates of teacher effects
 - We care about where in the effectiveness distribution teachers fall
 - VAM estimates can be wrong, but not so wrong that they radically change the estimated teacher-effectiveness distribution
 - We don't know much about how or whether VAM errors influence where teachers fall in the distribution
- Are we holding VAMs to a higher standard?
 - Estimates of productivity may be as imprecise and vary as much in the private sector

Focus of this Work

Assess the stability of (value-added) teacher job performance estimates over time, with a focus on pre- and post-tenure

North Carolina Data

- Administrative records for all NC teachers and students for grades 3-8 from 1995-96 to 2005-06
 - Fifth-grade performance for students with full history of test scores & in classes with 10-29 students
- Track teachers for whom we observe for at least two years pre-tenure and one year post-tenure
 - 281 unique teachers in this select sample

Analytic Approach

- $A_{i,j,t,s,g=5} = \alpha \mathbf{A}_{i(\text{history})} + X_{i,t,g=5} \gamma + \tau_{j,t,g=5} + \varepsilon_{i,j,t,s,g=5}$ where $\mathbf{A}_{i(\text{history})} = [A_{i,R,g=4} | A_{i,M,g=4} | A_{i,R,g=3} | A_{i,M,g=3}]$
- Specification is consistent with the unbiased estimates from Kane and Staiger (2008) and the bias-minimizing specification in Rothstein (2008)

Intertemporal Stability

- Variation between teachers explains 52 percent of overall variation in reading and 63 percent in math
 - Inclusion of time-varying teacher variables has only a trivial impact on the above distribution
- Average correlation of teacher job performance is 0.32 in reading and 0.54 in math
 - Estimates of stability of job performance are not terribly different from private sector estimates
- No evidence of convergence of teacher job performance as teachers progress through their careers
 - Variance of estimated teacher effects appears unrelated to tenure status or experience groupings

Pre- and Post-Tenure Job Performance Rankings: Reading

<i>Panel A. Using first two years of performance to predict post-tenure performance</i>						
	Post-tenure Quintile Rank					
Pre-tenure Quintile Rank	Bottom Quintile	Second Quintile	Third Quintile	Fourth Quintile	Top Quintile	Total Teachers
Bottom Quintile	32%	23%	19%	16%	11%	57
Second Quintile	27%	14%	27%	18%	14%	56
Third Quintile	21%	23%	30%	18%	7%	56
Fourth Quintile	16%	27%	18%	18%	21%	56
Top Quintile	5%	13%	5%	30%	46%	56
Total Teachers	57	56	56	56	56	281

<i>Panel B. Using first three years of performance to predict post-tenure performance</i>						
	Post-tenure Quintile Rank					
Pre-tenure Quintile Rank	Bottom Quintile	Second Quintile	Third Quintile	Fourth Quintile	Top Quintile	Total Teachers
Bottom Quintile	26%	30%	18%	14%	12%	50
Second Quintile	28%	14%	38%	12%	8%	50
Third Quintile	26%	24%	16%	22%	12%	50
Fourth Quintile	12%	18%	22%	24%	24%	50
Top Quintile	8%	14%	6%	28%	44%	50
Total Teachers	50	50	50	50	50	250

Pre- and Post-Tenure Job Performance Rankings: Math

<i>Panel A. Using first two years of performance to predict post-tenure performance</i>						
	Post-tenure Quintile Rank					
Pre-tenure Quintile Rank	Bottom Quintile	Second Quintile	Third Quintile	Fourth Quintile	Top Quintile	Total Teachers
Bottom Quintile	44%	25%	14%	16%	2%	57
Second Quintile	25%	30%	25%	13%	7%	56
Third Quintile	14%	14%	30%	18%	23%	56
Fourth Quintile	14%	18%	18%	23%	27%	56
Top Quintile	4%	13%	13%	30%	41%	56
Total Teachers	57	56	56	56	56	281

<i>Panel B. Using first three years of performance to predict post-tenure performance</i>						
	Post-tenure Quintile Rank					
Pre-tenure Quintile Rank	Bottom Quintile	Second Quintile	Third Quintile	Fourth Quintile	Top Quintile	Total Teachers
Bottom Quintile	42%	26%	18%	10%	4%	50
Second Quintile	36%	28%	20%	12%	4%	50
Third Quintile	16%	24%	26%	18%	16%	50
Fourth Quintile	4%	14%	20%	28%	34%	50
Top Quintile	2%	8%	16%	32%	42%	50
Total Teachers	50	50	50	50	50	250

De-selecting Poor Performers in Either Subject

Pre-tenure performance

Reading

Top 20%
60-80%
40-60%
20-40%
Bottom 20%

Math

Top 20%
60-80%
40-60%
20-40%
Bottom 20%

83 out of 281 teachers in bottom quintile in **either** subject

Post-tenure performance

Reading

Top 20%	11%
60-80%	18%
40-60%	20%
20-40%	23%
Bottom 20%	28%

Math

Top 20%	8%
60-80%	16%
40-60%	18%
20-40%	23%
Bottom 20%	35%



De-selecting Poor Performers in Both Subjects

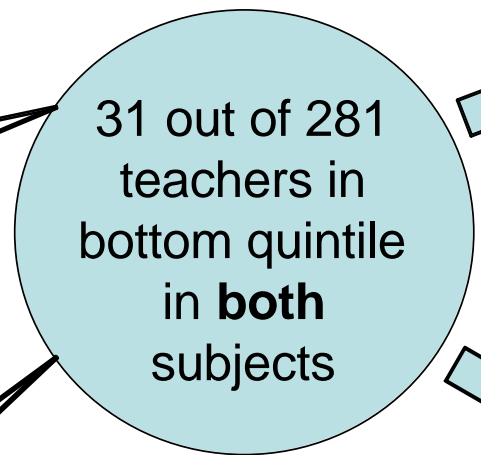
Pre-tenure performance

Reading

Top 20%
60-80%
40-60%
20-40%
Bottom 20%

Math

Top 20%
60-80%
40-60%
20-40%
Bottom 20%



Post-tenure performance

Reading

Top 20%	6%
60-80%	23%
40-60%	13%
20-40%	13%
Bottom 20%	45%

Math

Top 20%	0%
60-80%	13%
40-60%	19%
20-40%	19%
Bottom 20%	48%



In the Eye of the Beholder

- Year-to-year job performance estimates are modest (0.3 in reading and 0.5 in math); pre- and post-tenure estimates are somewhat higher (0.4 in reading and 0.6 in math)
 - We can't know whether these fluctuations represent true changes in job performance
- Inter-temporal estimates are not out of line with those found in other sectors of the economy that use them for policy purposes; and pre-tenure estimates clearly do predict estimated post-tenure performance
- How did we get here?
 - Poor evaluation/little use of evaluation today
 - Policymakers hope: VAMs are objective evaluation tool, which allows schools to do what they did not do when left to their own devices
- More research needed on using VAM to identify individual teacher effectiveness
 - Perfect can be the enemy of the good; we cannot learn all of what we need to know outside of actual policy variation

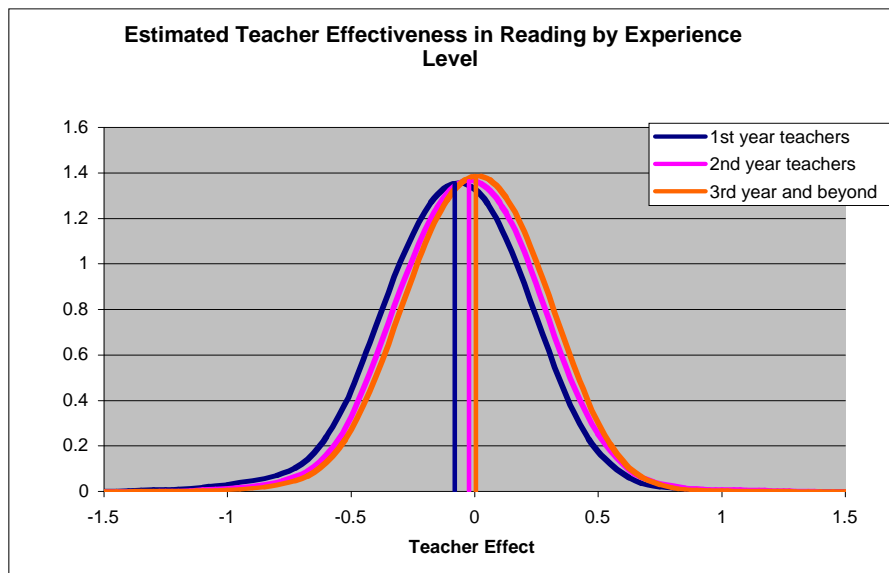
Hyperlink Slides

Comparison of Teacher Effects in Math by Passing Status

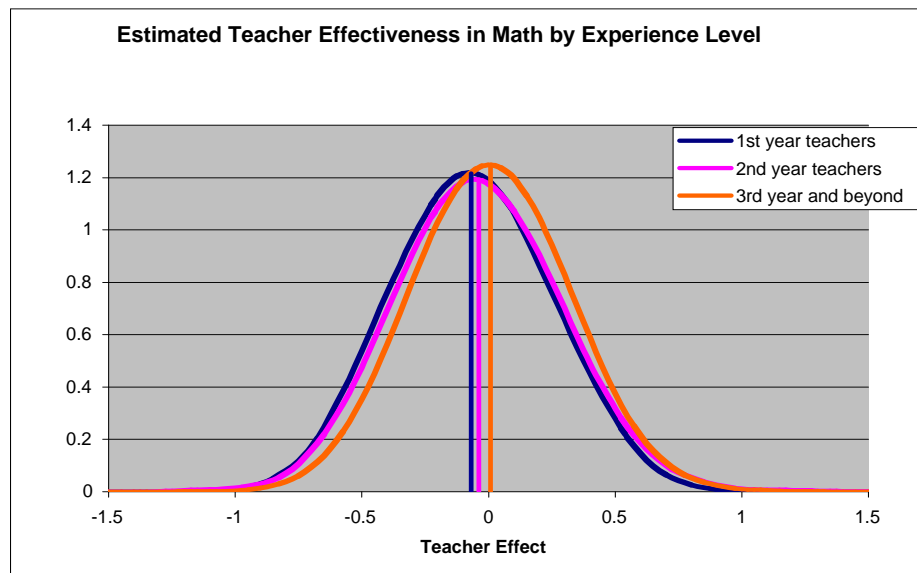
QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

[back](#)

Experience Levels



1st year mean-2nd year mean: 0.059** sd
2nd year mean-3rd year plus mean: 0.026* sd

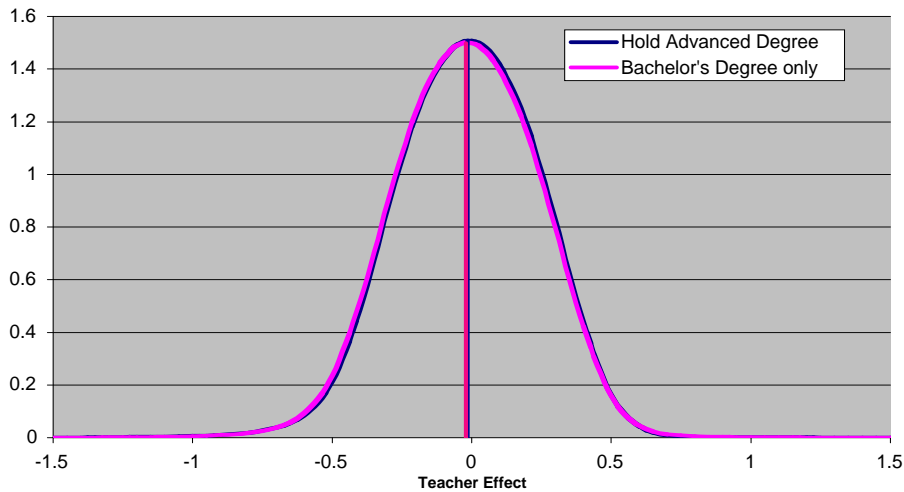


1st year mean-2nd year mean: 0.050* sd
2nd year mean-3rd year plus mean: 0.039** sd

[back degrees](#)

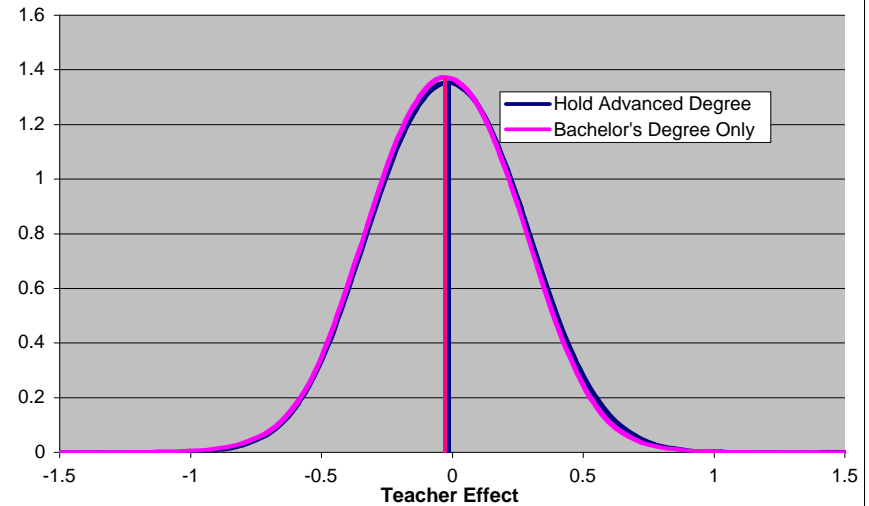
Degree Levels

Estimated Teacher Effectiveness in Reading by Degree Status



Difference in means: .005 sd

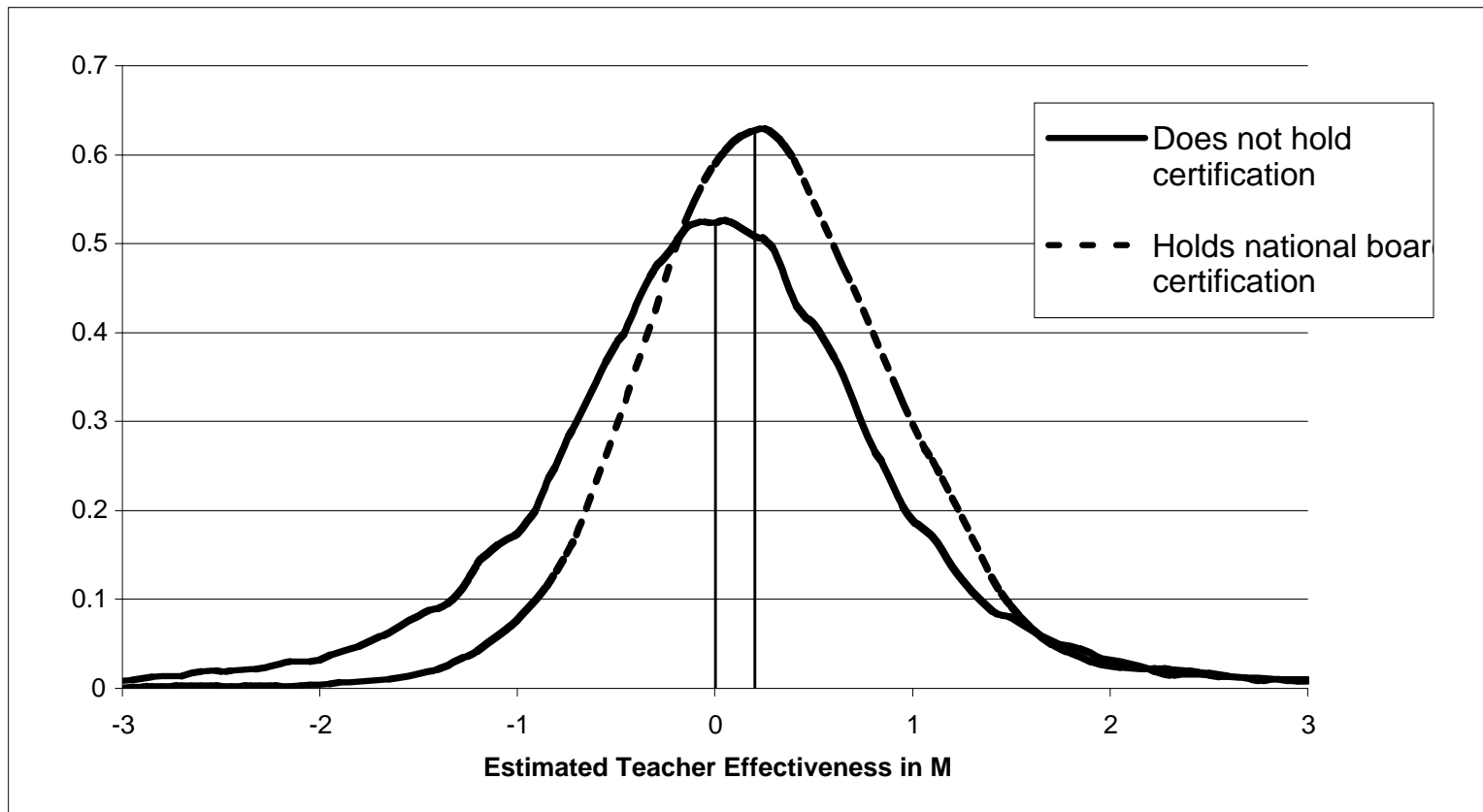
Estimated Teacher Effectiveness in Math by Degree Status



Difference in means: .014 sd

[back experience](#)

NBPTS Certification Status



Difference in means: 0.19** sd of teacher quality

[back degrees experience](#)