Instability of Teacher Effects Estimates from Value-Added Models

Edward Haertel

School of Education, Stanford University

CTC / PACE Symposium on

Linking Teaching with Learning: Examining Value Added Assessment Methodology for California

October 6, 2008

These preliminary findings are from work in progress, as reported at AERA (2008). I wish to acknowledge the contributions of my co-authors:

- Linda Darling-Hammond
  Stanford University
- Xiaoxia Newton
  University of California at Berkeley
- Ewart Thomas
  Stanford University

- Teachers for a New Era (TNE) longitudinal study offered opportunity to look closely at stability of teacher effects estimates
  - Using alternative statistical models
  - Teaching different courses
  - From year to year

Small Sample, Rich Data—TNE Study Includes:

- STEP and non-STEP teachers of HS Math and ELA
- Longitudinal data (full data for 2005-06 and 2006-07, as well as prior-year achievement)
- Variables describing
  - teacher preservice preparation
  - teaching assignments
  - student characteristics
  - student outcomes

Table 1: List of Sample for Math and ELA for the VAM Analysis

Academic Year Sample		2005-06	2006-07
Math teacher-course combinations [a]		57	46
ELA teacher-course combinations [b]		51	63
Students	Grade 9	646	881
	Grade 10	714	693
	Grade 11	511	789

[a] Some teachers taught multiple courses. There were 13 such math teachers for year 2005-06 and 10 for year 2006-07.

[b] The numbers of ELA teachers who taught multiple English courses for the two years were 16 and 15 respectively.

Our “Value-Added” Models

- OLS regressions assessing teachers’ value-added student achievement controlling for:
  - Model 1: students’ prior achievement only
  - Model 2: prior achievement plus demographics
  - Model 3: prior achievement plus school (as fixed effect)
  - Model 4: prior achievement plus demographics plus school (as fixed effect)

Findings

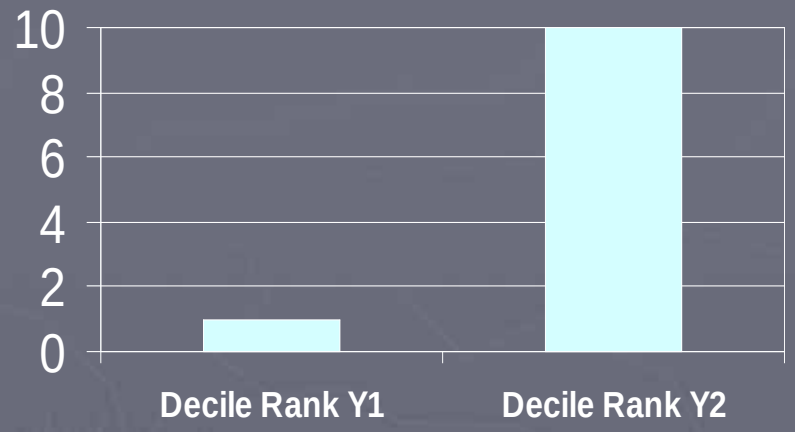
- Teachers' effectiveness rankings vary considerably, according to:
  - The statistical model used
  - The course taught
  - The year measured

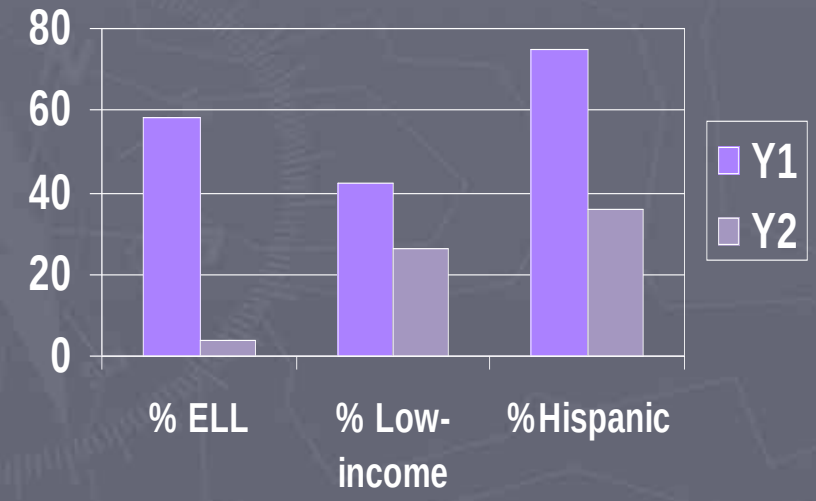
% of Teachers Whose Effectiveness Ratings Change

	By at least 1 decile	By at least 2 deciles	By at least 3 deciles
Across models*	56-80%	12-33%	0-14%
Across courses*	85-100%	54-92%	39-54%
Across years*	74-93%	45-63%	19-41%

*Depending on the model

One Extreme Case: An English language arts teacher





- Comprehensive high school
- Not a beginning teacher
- White
- Teaching English I
- Estimate controls for:
  - Prior achievement
  - Demographics
  - School fixed effect

What Have We Learned?

- Measures of teacher effects are not highly stable, and may depend on
  - Student backgrounds
  - School contexts
  - Different courses (as these intersect with teacher skills and / or student backgrounds?)
  - Different years (as these represent experience, context, course, or student differences?)
- More research is needed to ascertain what we are actually measuring when we seek to measure a “teacher effect”
- Validity of VAM teacher effect estimates for high stakes inferences is not yet established

Differences across Models

Differences in teachers' ratings across models are significantly correlated with the characteristics of their students, including:

- Free / reduced price lunch status
- English language learner status
- Race / ethnicity (proportion of African American, Asian, Latino students)
- Parent education
- "On track" status for math course taken

Differences Across Courses

Individual teachers are rated differently when teaching different classes.

- For most models, ratings of effectiveness for the same teacher across two courses are not significantly correlated
- In 16 cases where several teachers were teaching the same 2 different courses, ANOVA (with prior test score controls) found that the course was more often a significant predictor of student achievement (11 cases) than the teacher (3 cases).

Differences Across Years

- While correlations of teacher ratings were significantly correlated across years (r = .3 to .4 across models in ELA and .4 to .6 in math), there were still noticeable changes in teachers “effectiveness” scores from one year to the next.

Thank you