

# Teacher Compensation & the Level and Distribution of Teaching Quality

Bruce D. Baker  
Rutgers University

12/18/2010

1

Bruce D. Baker © 2010

## Overarching Policy Objectives

- Improving Average Teacher Quality
- Distributing Teacher Quality Fairly

12/18/2010

2

Bruce D. Baker © 2010

## Quick Note on Proposals for Redistributing Quality

- Some (NCTQ) believe that inequities in teacher quality are largely a within-district problem, where poor schools in districts get the weakest teachers and rich schools in the same district get the best teachers.
  - While such problems do exist, this argument ignores the reality that most districts don't have rich schools and poor schools. Some have poor and very poor schools, and others have rich and very rich schools.
- Therefore, the primary fix for TQ disparities lies in district level teacher contracts and with provisions like "seniority bumping."
  - The fix, in their view, is to decentralizing hiring to principals, eliminate seniority preferences and include "mutual consent" language in contracts.
  - They also believe this fix can be imposed through federal Title I restrictions

12/18/2010

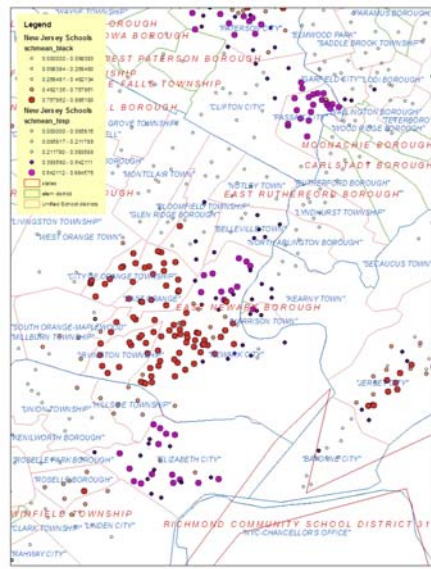
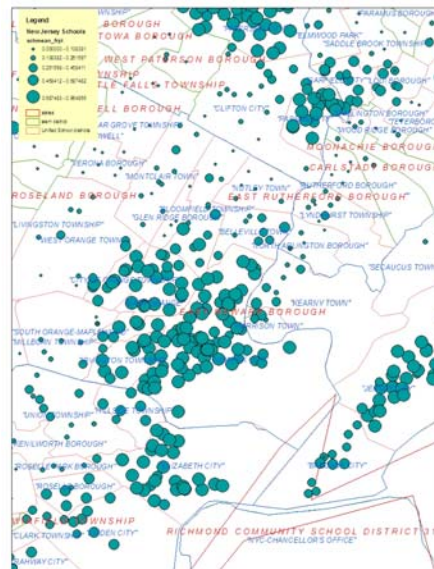
3

Bruce D. Baker © 2010

### Student Demography of Newark City Schools and Surroundings: Race & Poverty

% Free or Reduced Lunch

% Black or Hispanic



Data Source: NCES Common Core, Public School Universe 2007-08

Bruce D. Baker © 2010

## Findings of related research

- We conduct an interrupted time-series analysis of data from 1998-2005 and find that the shift from a seniority-based hiring system to a “mutual consent” hiring system leads to an initial increase in both teacher turnover and share of inexperienced teachers, especially in the district’s most disadvantaged schools. For the most part, however, these initial shocks are corrected within four years leaving little change in the distribution of inexperienced teachers or levels of turnover across schools of different advantage.
  - Gross & Goldhaber, 2010
  - [http://www.nctq.org/docs/Mutual\\_Concent\\_8049.pdf](http://www.nctq.org/docs/Mutual_Concent_8049.pdf)

12/18/2010

5

Bruce D. Baker © 2010

## Proposals for Improving Overall Quality

- Goal #1: Incentivize “good teaching” among current crop of teachers
  - Most studies find these approaches generally ineffective (when applied at individual level as fiscal incentive)
- **Goal #2: Alter incentive/compensation structure of profession to attract stronger candidates to teaching**
  - Level of pay/overall compensation package
  - Risk/reward ratio
  - Incentives/compensation structure

12/18/2010

6

Bruce D. Baker © 2010

## Research-Policy Sleight of Hand

- “For example, it’s clear from abundant research that paying teachers only on the basis of their degrees and years of experience is not in the best interest of students or teachers. As the National Council on Teacher Quality, a research and policy organization whose board of directors I chaired for several years, put it, “the evidence is conclusive that master’s degrees do not make teachers more effective.”
  - Andrew Rotherham in Time Magazine
- Logical conclusion - we should prohibit outright any compensation being based on masters degrees or on experience & should use the savings (from taking away all of that money from those who already have it?) to pay for easily measurable teaching quality! (an obvious revenue-neutral path to dramatic improvements!)

12/18/2010

7

Bruce D. Baker © 2010

## The question not asked

- What has been studied
  - Do teachers who hold general masters degrees, versus those who do not, scattered across a variety of settings, show differences in the average outcome gains of their students?
  - Do teachers at varied levels of experience, scattered across a variety of settings, show differences in the average outcome gains of their students?
- **What we do not know!**
  - Studies of the association between different levels of experience and the association between having a masters degree or not and student achievement gains have never attempted to ask about the potential labor market consequences of stopping providing additional compensation for teachers choosing to further their education – even if only for personal interest – or stopping providing any guarantee that a teacher’s compensation will grow at a predictable rate over time throughout the teacher’s career.
  - The adverse labor market effects may be particularly strong if we replace predictable salary increments (however frustrating) with very noisy performance measures significantly outside control of teachers.

12/18/2010

8

Bruce D. Baker © 2010

# Redesigning Compensation

A few examples from Urban Districts  
Or... *is there a single definition of  
the “status quo?”*

12/18/2010

9

Bruce D. Baker © 2010

## Two Existing District Frameworks

- Denver ProComp (part of the status quo?)
  - Incentives to work in high need schools
  - Knowledge and Skill Based Pay
    - Odden and Kelley (2001)
    - Tiered system of professional review
    - Includes share based on student assessment (<20%)
    - Retains experience & degree level pay
    - See: <http://denverprocomp.dpsk12.org/>
- See Also: [http://www.cpre.org/images/stories/cpre\\_pdfs/rr50.pdf](http://www.cpre.org/images/stories/cpre_pdfs/rr50.pdf)

12/18/2010

10

Bruce D. Baker © 2010

## Two Existing District Frameworks

- Washington DC IMPACT (Current reform model)
  - Type 1 Teachers
    - Individual Value-Added Student Achievement Data (IVA) — This is a measure of the impact you have on your students' learning over the course of the school year, as evidenced by the DC CAS. **This component makes up 50% of your IMPACT score.**
    - Teaching and Learning Framework (TLF) — This is a measure of your instructional expertise. This component makes up 35% of your IMPACT score.
    - Commitment to the School Community (CSC) — This is a measure of the extent to which you support and collaborate with your school community. This component makes up 10% of your IMPACT score.
    - School Value-Added Student Achievement Data (SVA) — This is a measure of the impact your school has on student learning over the course of the school year, as evidenced by the DC CAS. This component makes up 5% of your IMPACT score.
    - Core Professionalism (CP) — This is a measure of four basic professional requirements for all school-based personnel. This component is scored differently from the others, which is why it is not represented in the pie chart. For more information, please see the Core Professionalism section of this guidebook.

12/18/2010

11

Bruce D. Baker © 2010

What do we know about  
average salaries and teacher  
quality?  
(and related economic  
constraints)

12/18/2010

12

Bruce D. Baker © 2010

## What do we know about Average Teacher Quality & Salaries?

- Murnane and Olson (1989) find that salaries affect the decision to enter teaching and the duration of the teaching career.<sup>[1]</sup>
- Figlio (1997, 2002) and Ferguson (1991) find that higher salaries are associated with better qualified teachers<sup>[2]</sup>
- Loeb and Page (1998, 2000) find that raising teacher wages by ten percent reduces high school dropout rates by between three and six percent and increases college enrollment rates by two percent.<sup>[3]</sup>

<sup>[1]</sup> Richard J. Murnane and Randall Olsen (1989) The effects of salaries and opportunity costs on length of state in teaching. Evidence from Michigan. *Review of Economics and Statistics* 71 (2) 347-352

<sup>[2]</sup> David N. Figlio (1997) Teacher Salaries and Teacher Quality. *Economics Letters* 55 267-271. David N. Figlio (2002) Can Public Schools Buy Better-Qualified Teachers? *Industrial and Labor Relations Review* 55, 686-699. Ronald Ferguson (1991) Paying for Public Education: New Evidence on How and Why Money Matters. *Harvard Journal on Legislation*. 28 (2) 465-498.

<sup>[3]</sup> Susanna Loeb and Marianne Page (2000) Examining the link between teacher wages and student outcomes: the importance of alternative labor market opportunities and non-pecuniary variation. *Review of Economics and Statistics* 82, 393-408. Susanna Loeb and Marianne Page (1998) Examining the link between wages and quality in the teacher workforce. Department of Economics, University of California, Davis.

## External Economic Constraints & Teacher Quality

- David N. Figlio and Kim S. Rueben
  - Tax limits and the qualifications of new teachers
  - [Journal of Public Economics Volume 80, Issue 1](#), April 2001, Pages 49-71
    - This paper examines the impact of local tax limits on new teacher quality. Using data from the National Center for Education Statistics we find that tax limits systematically reduce the average quality of education majors, as well as new public school teachers in states that have passed these limits. The average relative test scores of education majors in tax limit states declined by ten percent as compared to the relative test scores of education majors in states that did not pass limits. This relationship is strengthened if we control for school finance equalization reforms or examine tax limits passed in two different periods.

# Can we measure Teacher Effectiveness with Student Assessment Data?

12/18/2010

15

Bruce D. Baker © 2010

## Section I

What is VAM?

12/18/2010

16

Bruce D. Baker © 2010



## Intent of VAM

- Value-added Modeling
- To isolate/identify/estimate the relationship between having teacher A or teacher B on the average achievement gains of students with each teacher
- VAM is more complex than simply taking the average difference of test scores at time T+1 minus test scores at time T.

12/18/2010

17

Bruce D. Baker © 2010

## Basic Assumption

Teacher Effectiveness = Student Test  
Scores After - Student Test Scores Before

(“controlling” for a variety of conditions)

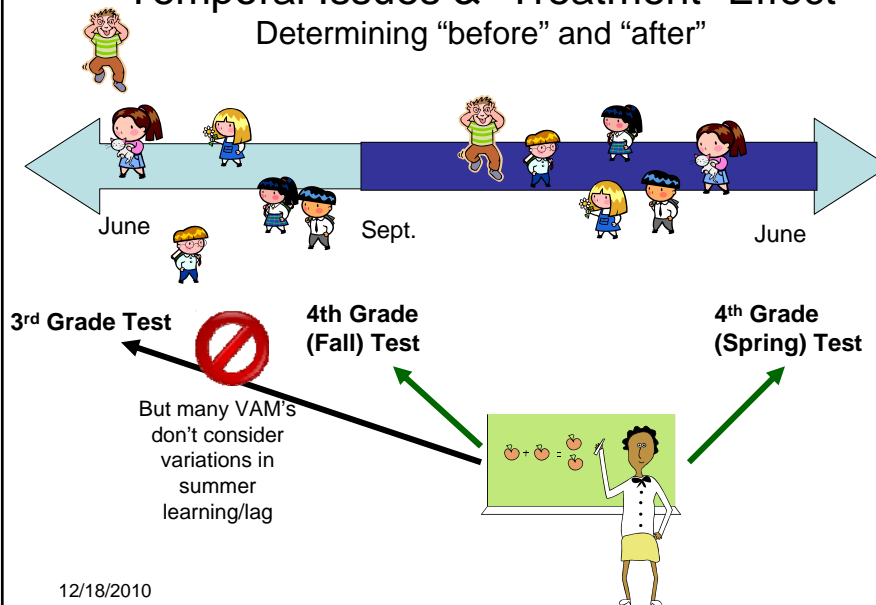
12/18/2010

18

Bruce D. Baker © 2010

## Temporal Issues & “Treatment” Effect

### Determining “before” and “after”

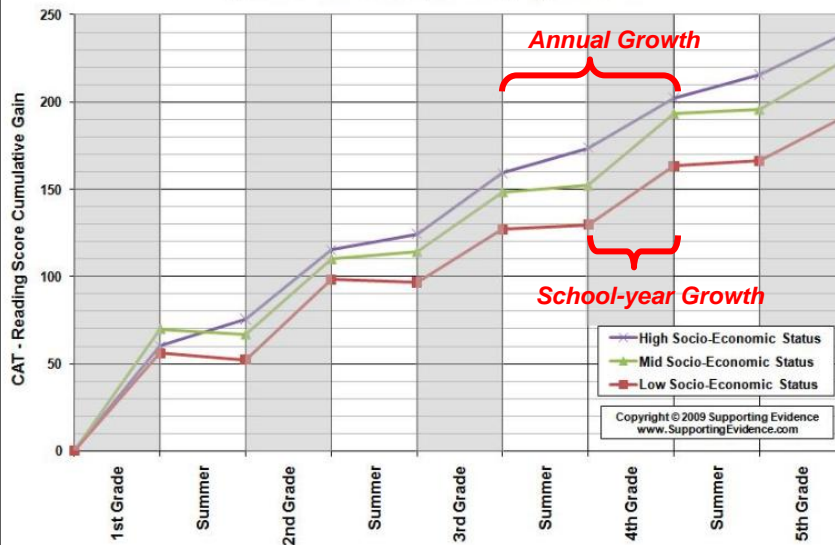


19

Bruce D. Baker © 2010

## Summer Learning Lag by Poverty

CAT - Reading Score Cumulative Gains From Grades 1 to 5  
By Socio-Economic Status, Baltimore, MD, 1982 - 1987



<http://epa.sagepub.com/content/23/2/171.abstract>

Bruce D. Baker © 2010

## New Gates Study Finding

- The norm sample results imply that students improve their reading comprehension scores just as much (or more) between April and October as between October and April in the following grade. Scores may be rising as kids mature and get more practice outside of school.

12/18/2010

21

Bruce D. Baker © 2010

## Issues with using VAM to evaluate individual teacher “effectiveness”

- Statistical
  - Measurement
    - Noise/error rate (instrument noise/error)
      - Inter-temporal variation
    - Scale
    - Test/Form
  - Application
    - Unexplainable variation (contextualized noise)
    - Non-random assignment
      - School level issues
      - District level issues/neighborhood
      - State level issues (segregation)

12/18/2010

22

Bruce D. Baker © 2010

## Notes on Stability of Ratings

- ONLY “About one quarter to one third of the teachers in the bottom and top quintiles stay in the same quintile from one year to the next while roughly 10 to 15 percent of teachers move all the way from the bottom quintile to the top and an equal proportion fall from the top quintile to the lowest quintile in the next year.” (p. 2)<sup>[1]</sup>

<sup>[1]</sup> Sass, T.R. (2008) The Stability of Value-Added Measures of Teacher Quality and Implications for Teacher Compensation Policy. Urban Institute, [http://www.urban.org/UploadedPDF/1001266\\_stabilityofvalue.pdf](http://www.urban.org/UploadedPDF/1001266_stabilityofvalue.pdf) See also: McCaffrey, Daniel F.; Tim R. Sass; J. R. Lockwood and Kata Mihaly. 2009. “The Intertemporal Variability of Teacher Effect Estimates.” *Education Finance and Policy*, 4(4), pp. 572-606.

12/18/2010

23

Bruce D. Baker © 2010

## New Gates Findings

VARIABLE	DIFFERENT SECTION			PRIOR YEAR		
	Total Variance	Correlation	Stable Component	Total Variance	Correlation	Stable Component
Test						
State Math Test	0.053	0.380	0.020	0.040	0.404	0.016
	[0.231]		[0.143]	[0.20]		[0.127]
State ELA Test	0.032	<b>0.179</b>	0.006	0.028	<b>0.195</b>	0.005
	[0.178]		[0.075]	[0.166]		[0.073]
BAM Test	0.071	<b>0.227</b>	0.016			
	[0.266]		[0.127]			
Stanford 9 OE ELA	0.129	0.348	0.045			
	[0.359]		[0.212]			

12/18/2010

24

Bruce D. Baker © 2010

# New Gates Finding

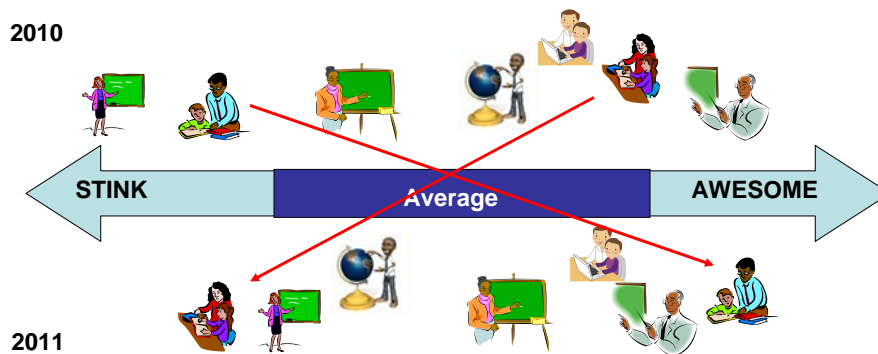
- When the between-section or between-year correlation in teacher value-added is below .5, the implication is that more than half of the observed variation is due to transitory effects rather than stable differences between teachers. That is the case for all of the measures of value-added we calculated.
- Authors note that given that overall variance is large, having teacher effects explain even a small share of that variance is important. Authors estimate “stable component of variance.”

12/18/2010

25

Bruce D. Baker © 2010

# Stability of Ratings



12/18/2010

26

Bruce D. Baker © 2010

## Basing tenure on sequential VAM success...

- Many have discussed the idea that teachers should not be granted tenure unless they can string together 3 consecutive years of successful VAM ratings
- Teachers in their first two years have a hard time getting a positive rating
- It may take several years after that to get lucky enough to string together 3 “good” years. And yes, I do mean lucky!
- For any given entering cohort of 100 teachers, we don’t know how many would even be tenurable after 10, or even 15 years.

12/18/2010

27

Bruce D. Baker © 2010

## Notes on Misidentification Due to “random error”

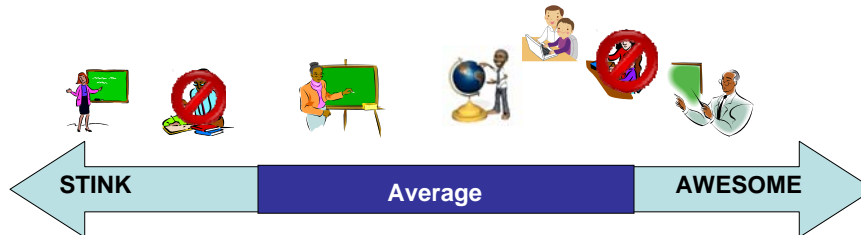
- There is about a 25% chance, if using three years of data or 35% chance if using 1 year of data that a teacher who is “average” would be identified as “significantly worse than average” and potentially be fired
- Of particular concern is the likelihood that a “good teacher” is falsely identified as a “bad” teacher, in this case a “false positive” identification.
  - According to the study, this occurs 1 in 10 times (given three years of data and 2 in 10 given only one year).
  - Also problematic from a policy perspective but perhaps less so from a legal perspective - because it results in improper retention rather than improper dismissal - is the equal likelihood of a “false negative error,” that a “bad teacher” is improperly identified as a “good one.”

Schochet, Peter Z. and Hanley S. Chiang (2010). *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains* (NCEE 2010-4004). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

28

Bruce D. Baker © 2010

# Classification Error



12/18/2010

29

Bruce D. Baker © 2010

## Different Tests Yield Different Ratings!

- Sean Corcoran (2010) explains that “Houston has administered two standardized tests every year: the state TAKS and the nationally normed Stanford Achievement Test.”
- “among those who ranked in the top category (5) on the TAKS reading test, more than 17 percent ranked among the lowest two categories on the Stanford test. Similarly, more than 15 percent of the lowest value-added teachers on the TAKS were in the highest two categories on the Stanford.”

Corcoran, Sean P., Jennifer L. Jennings, and Andrew A. Beveridge. 2010. “Teacher Effectiveness on High- and Low-Stakes Tests.” Paper presented at the Institute for Research on Poverty summer workshop, Madison, WI.

12/18/2010

30

Bruce D. Baker © 2010

## Related Gates Finding

- the correlation between a teacher's value-added on the state test and their value-added on the Balanced Assessment in Math was .377 in the same section and .161 between sections.
- we estimate the correlation between the *persistent component of teacher impacts* on the state test and on BAM is moderately large, .54.
- The correlation in the *stable teacher component* of ELA value-added and the Stanford 9 OE was lower, .37.

12/18/2010

31

Bruce D. Baker © 2010

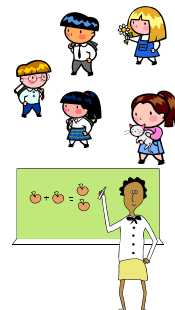
## Non-random Assignment School Level



Principal



Teacher A



Teacher B

12/18/2010

32

Bruce D. Baker © 2010



## Potential reasons for school-level non-random assignment

- Finding the “best match” for each child
- A teacher’s desire to try to help out the most difficult kids
  - VAM would stamp this out!
- A principal’s desire to make a teacher’s life difficult (and perhaps even get that teacher fired for low VA scores)
- Most interested/aggressive parents requesting a specific teacher

12/18/2010

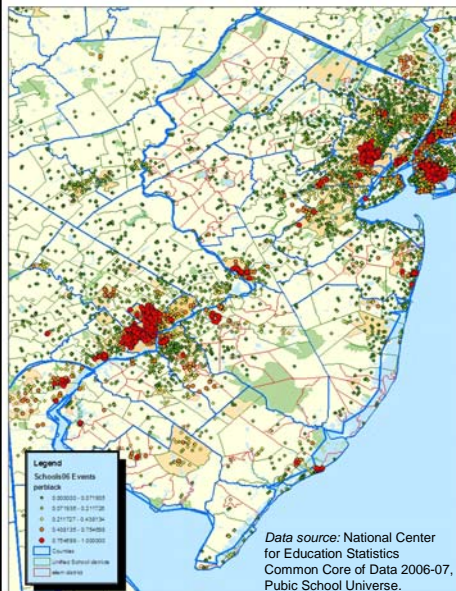
33

BRUCE D. EMMER © 2010

## Non-random assignment statewide!

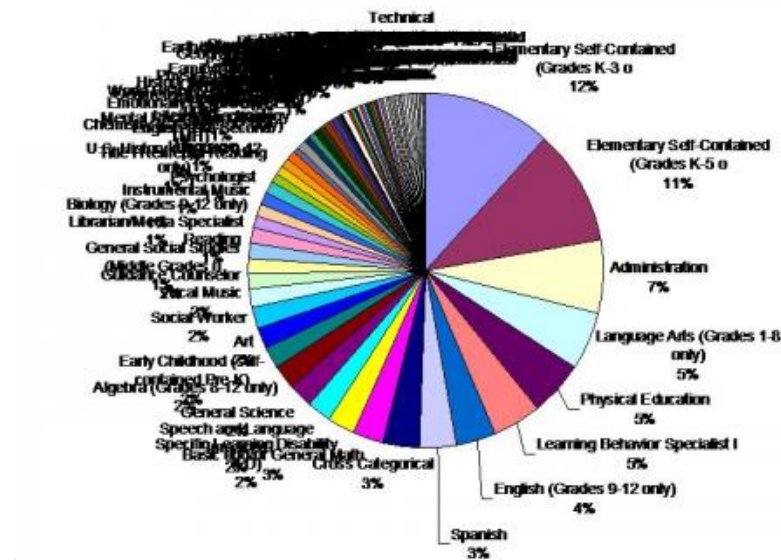
% Black

% Hispanic



BRUCE D. EMMER © 2010

## Illinois Staffing File Data



1

35

Bruce D. Baker © 2010

## Issues Cont'd

- Writing a separate contract for the <20% of teachers who can be attached to math/reading tests
  - Those who can be directly linked (with prior scores available) to student test scores in reading & math in grades 3 to 8
- Isolating teacher effect over other effects
  - Other teacher's effects: Spillover
  - Non-random assignment (clustering)
    - Unmeasured student characteristics
    - Collective effects (peer)

12/18/2010

36

Bruce D. Baker © 2010

## Spillover Effects

- Bruegmann (2009), for example, found in a study of North Carolina teachers that students perform better, on average, when their teachers have more effective colleagues.[\[1\]](#)
- Koedel (2009) found that reading achievement in high school is influenced by both English and math teachers.[\[2\]](#)

[\[1\]](#) Jackson, C. Kirabo, and Elias Bruegmann. 2009. "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers," *American Economic Journal: Applied Economics* 1:85–108

[\[2\]](#) Koedel, Cory. 2009. "An empirical analysis of teacher spillover effects in secondary school," *Economics of Education Review* 28:682–692.

12/18/2010

37

Bruce D. Baker © 2010

## Issues Cont'd

- Finally, after creating these adverse work conditions for that 20%, finding "better" teachers to replace the ones you wrongly fire.

Who will be waiting  
in line?



12/18/2010

38

Bruce D. Baker © 2010

# Some Up-side?

Principal observations of teachers  
linked to student outcomes?

12/18/2010

39

Bruce D. Baker © 2010

## Principal perceptions & Tchr. VA Ratings

- we examine the results of a randomized pilot program in which school principals were provided with estimates of the performance of individual teachers in raising their students' test scores in math and English.
  - objective teacher performance estimates based on student data and principals' prior beliefs are positively correlated, and the strength of this relationship rises with the precision of the objective estimates and the precision of subjective priors.
  - Second, principals who are provided with objective performance data incorporate this information into their posterior beliefs, and do so to a greater extent when the data are more precise and when their priors are less precise.
  - Rockoff, Staiger, Kane, Taylor, July 2010
  - <http://www.nber.org/papers/w16240>

12/18/2010

40

Bruce D. Baker © 2010

## Subjective Evals. & Tchr. VA Ratings

- Using data from New York City public schools, we estimate whether subjective evaluations of teacher effectiveness have predictive power for the achievement gains made by teachers' future students.
- We find that these subjective evaluations have substantial power, comparable with and complementary to objective measures of teacher effectiveness taken from a teacher's first year in the classroom.

Rockoff & Speroni, April 2010

[http://www0.gsb.columbia.edu/faculty/jrockoff/papers/rockoff\\_speroni\\_subjective\\_evals\\_AEA\\_PP\\_final.pdf](http://www0.gsb.columbia.edu/faculty/jrockoff/papers/rockoff_speroni_subjective_evals_AEA_PP_final.pdf)

12/18/2010

41

Bruce D. Baker © 2010

## Gates Study Student Perceptions

- In general, the student perception measures were highly correlated between sections taught by the same teacher.

12/18/2010

42

Bruce D. Baker © 2010

# Follow the Leader?

Which really outstanding states are leading the way with these teacher compensation reform strategies?

12/18/2010

43

Bruce D. Baker © 2010

## State Statutes

- Teacher evaluations must include at least 50% student test scores
  - Colorado
  - Louisiana
  - Tennessee
- Teacher evaluations must include between 33 and 50% test scores
  - Arizona
- Teacher evaluations must include some consideration of test scores
  - Connecticut
  - Michigan

12/18/2010

44

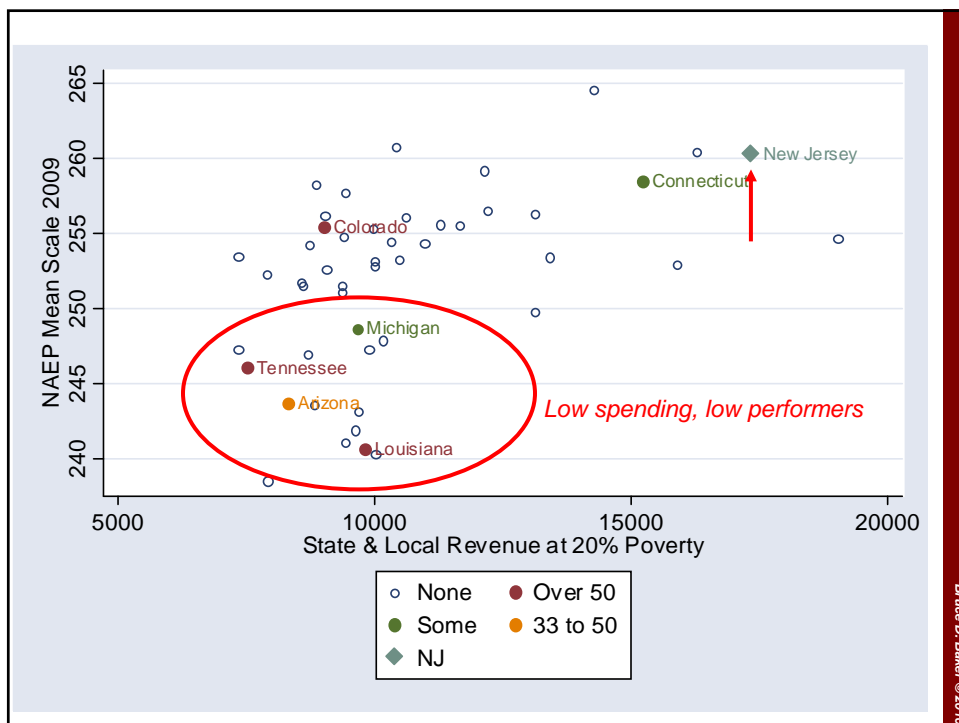
Bruce D. Baker © 2010

# Are these states really good education policy role models for New Jersey?

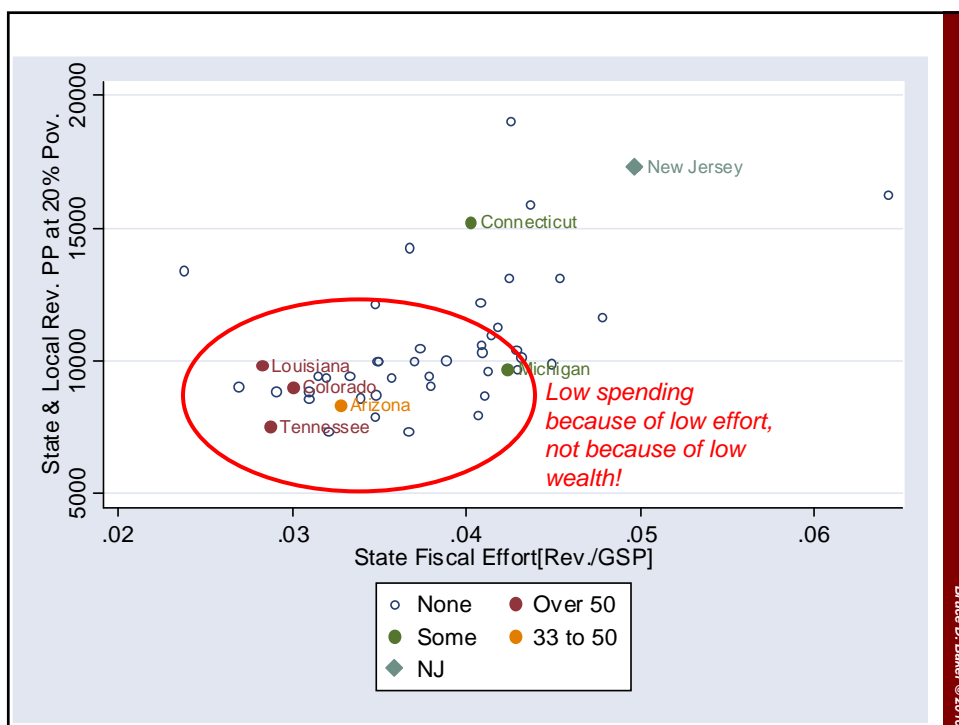
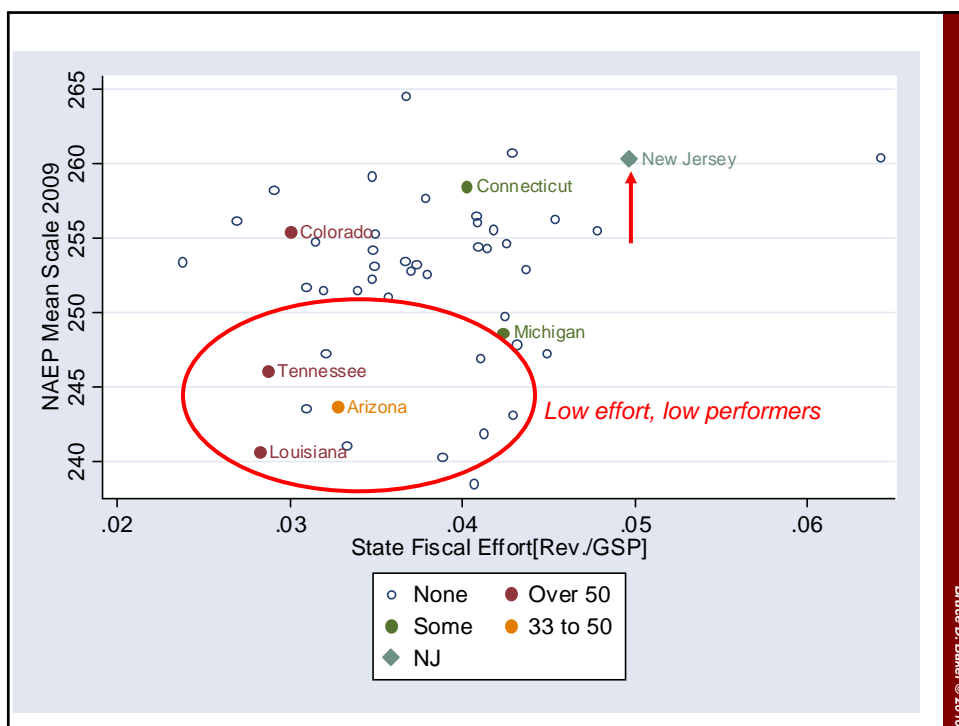
12/18/2010

45

Bruce D. Baker © 2010



Bruce D. Baker © 2010





## Policy Logic - But It's the Best Available Option????

If not “A” it must be “B”

12/18/2010

49

Bruce D. Baker © 2010

## Reformy Logic

- Something must be done
- This is something
- Therefore we must do it

12/18/2010

50

Bruce D. Baker © 2010

# Reformy Rule #1

**Anything > Status Quo**

12/18/2010

51

Bruce D. Baker © 2010

## Reformy Proof that VAM is better than Current Evaluations

- Because value-added modeling exists and purports to measure teacher effectiveness, it therefore counts as “something,” which is a subclass of “anything” and therefore it is better than the “status quo.” That is:

**Value-added modeling = “something”**

**Something (subset symbol) Anything (something is a subset of anything)**

**Something > Status Quo**

**Value-added modeling > Current Teacher Evaluation**

12/18/2010

52

Bruce D. Baker © 2010

## Additional proofiness

- After all, you can't even measure the error rate in current principal and supervisor evaluations of teachers can you? And if you can't measure the error rate it must be higher than any error rate you can measure?
- That is, the unobserved error rate in one system is necessarily greater than the observed error rate of another – even if we have no way to quantify it – in fact, because we have no way to quantify it?

**Unobserved error rate of 'status quo' >  
measured error rate of VAM**

12/18/2010

53

Bruce D. Baker © 2010

## Conclusion???

**Let's be really blunt here. Both are  
patently stupid arguments!**

12/18/2010

54

Bruce D. Baker © 2010

## Is “something” always better than “nothing”?

- If we were in a society that still walked pretty much everywhere, and some tech genius invented a new cool thing – called the automobile – but the automobile would burst into a superheated fireball on every fifth start, I think I’d keep walking until they worked out that little kink. If they never worked out that little kink, I’d probably still be walking.

12/18/2010

55

Bruce D. Baker © 2010



For now, I’d rather walk!



12/18/2010

56

Bruce D. Baker © 2010

## Comment

- It's not that Value-added measures are not or can not be useful...
  - They are extremely useful for evaluating/ monitoring systems at various levels of aggregation.
  - We need to continue developing our data and data systems to learn more about how these measures can reasonably inform system improvements.
- The problem lies in trying to pluck out any one point, or low-level of aggregation (teacher effect), from a large system and know with any level of certainty what that point represents.

12/18/2010

57

Bruce D. Baker © 2010

## Gates

- Assuming that the distribution of teacher effects is “bell-shaped” (that is, a normal distribution), this means that if one could accurately identify the subset of teachers with value-added in the top quartile, they would raise achievement for the average student in their class by .18 standard deviations relative to those assigned to the median teacher.<sup>6</sup> Similarly, the worst quarter of teachers would lower achievement by .18 standard deviations. So the difference in average student achievement between having a top or bottom quartile teacher would be .36 standard deviations. (p.19)

12/18/2010

58

Bruce D. Baker © 2010

# Gates Recommendations

- working with teachers to develop accurate lists of the students in their care, so that value-added data are as accurate as possible;
- using confidential surveys to collect student feedback on specific aspects of a teacher's practice (which could reach virtually every classroom, including those in non-tested grades and subjects);
- retraining principals and instructional coaches to do classroom observations in a more meaningful way; and
- delivering such data in a timely way to school principals and teachers.

12/18/2010

59

Bruce D. Baker © 2010

# Moving Forward

- We don't know enough about labor market impact - broad-based, or localized - to pick any one method and enforce top down reform
  - Value-added measures are far from ready for prime time use for rating individual teachers (but can be useful system diagnostic tools)
  - Incentive pay generally ineffective for changing behaviors in the system
- We must be cognizant of effects of our decisions on teacher pipeline and distribution
- Policy choices are not simple (and not dichotomous) -
  - "bad current policies vs. less bad VAM models" **WRONG!**
  - "compensation based on experience & degree level ineffective, therefore eliminating these factors is good!" **WE DON'T KNOW THAT!**
- Now is a good time to keep collecting (in NJ start collecting) good data and figuring out a) what we can learn from it and b) how to best use it!
- Now is also a good time to allow and encourage schools and districts, and teachers to negotiate creative compensation policies and to take the time to thoroughly evaluate those policies (require thorough evaluation!)

12/18/2010

60

Bruce D. Baker © 2010

## Other considerations

- Tiered systems of professional review (Denver ProComp)
  - These have been tried but we've done too little research on them
- Improved systems of peer evaluation and principal evaluation of teachers
  - Coupled with somewhat longer (perhaps) initial review period
  - Appropriate tightening of tenure requirements/raising the bar (while not ignoring replacement quality and/or churning/turnover)
- Getting higher need districts ahead of the game on hiring/timing
  - An issue on which my analyses and TNTP analyses of NPS agree!
- Taking a look at returns to experience on high need district salary schedules to ensure that they promote retention at appropriate career stages
  - In NJ, many high need district salary schedules are flat on front end, creating relative salary disadvantage both with respect to other professions and to other teaching positions.
- Better positioning high need districts to recruit “highly productive” teachers with experience/track record (based on great new study from Calder Center)

12/18/2010

61

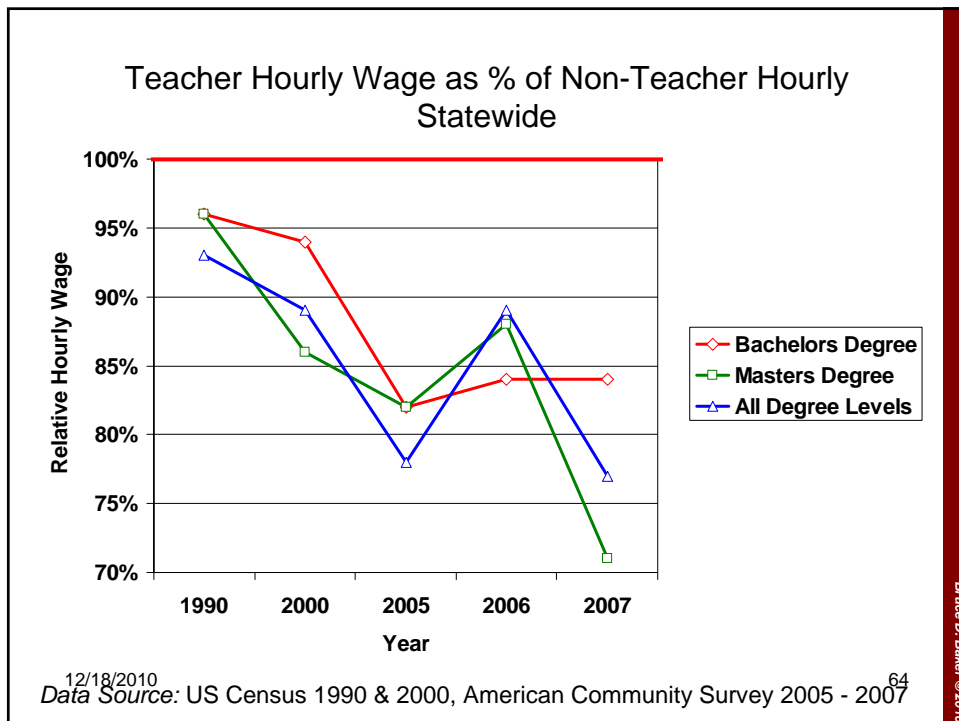
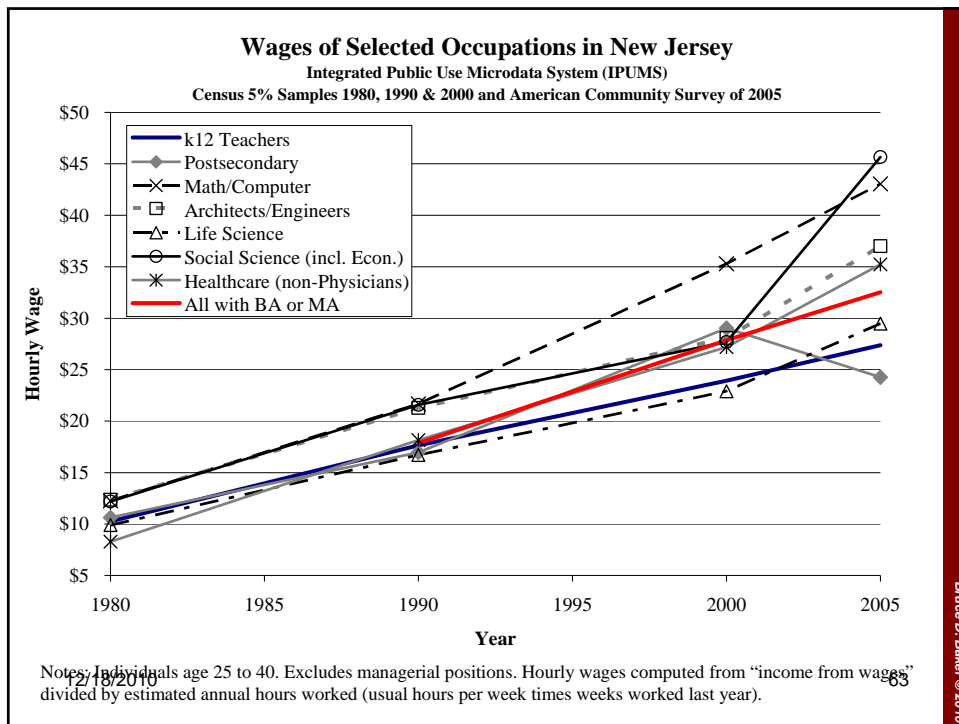
BRUCE D. BAKER © 2010

## What do we Know About NJ Teacher Wages

12/18/2010

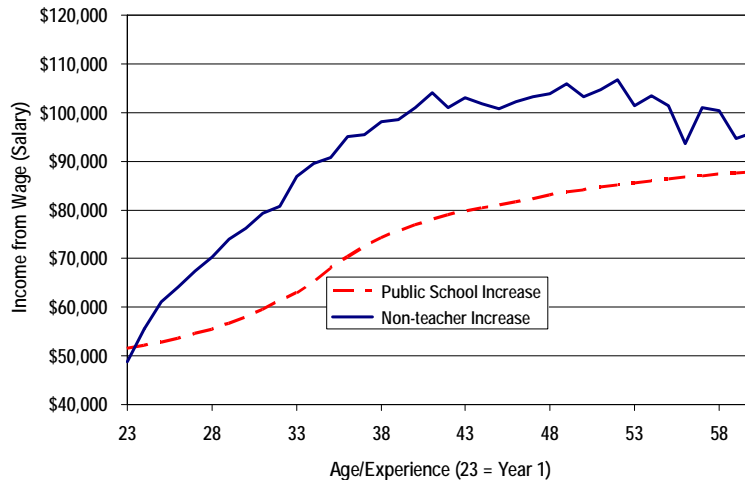
62

BRUCE D. BAKER © 2010





### Returns to Experience/Age for Teachers and Non-Teachers (at fixed degree level, location)



Data Source: Non-Teacher Wages from US Census 2000, American Community Survey 2005 - 2008 based on regression model of wages controlling for age, location, degree level and year. Teacher wages based on NJDOE Personnel Files also using regression model controlling for experience, degree level, location, position type and year.

Bruce D. Baker © 2010

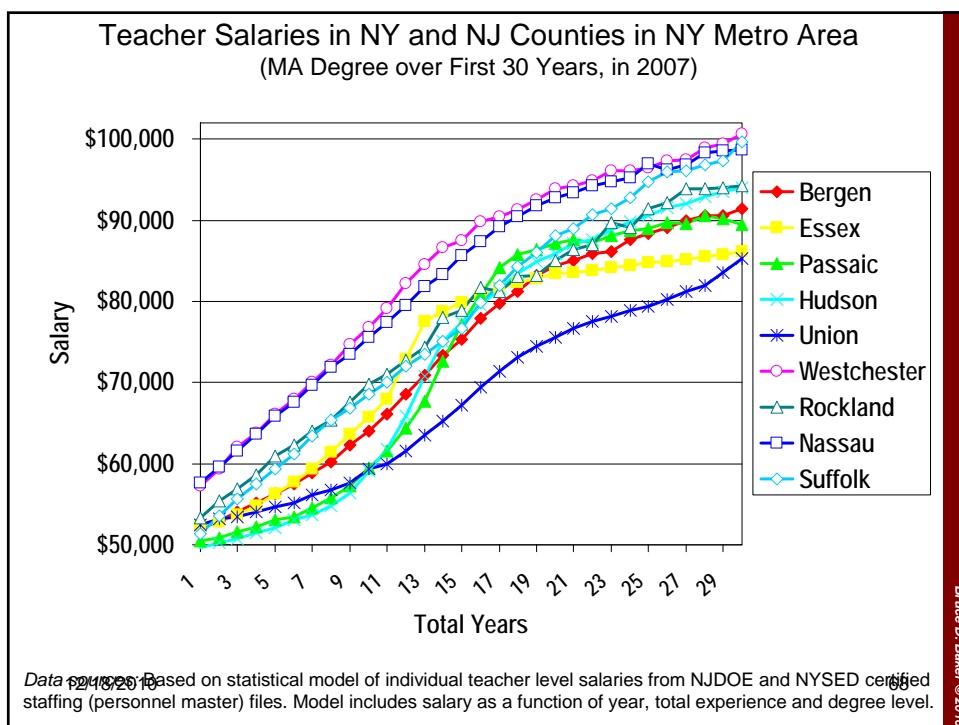
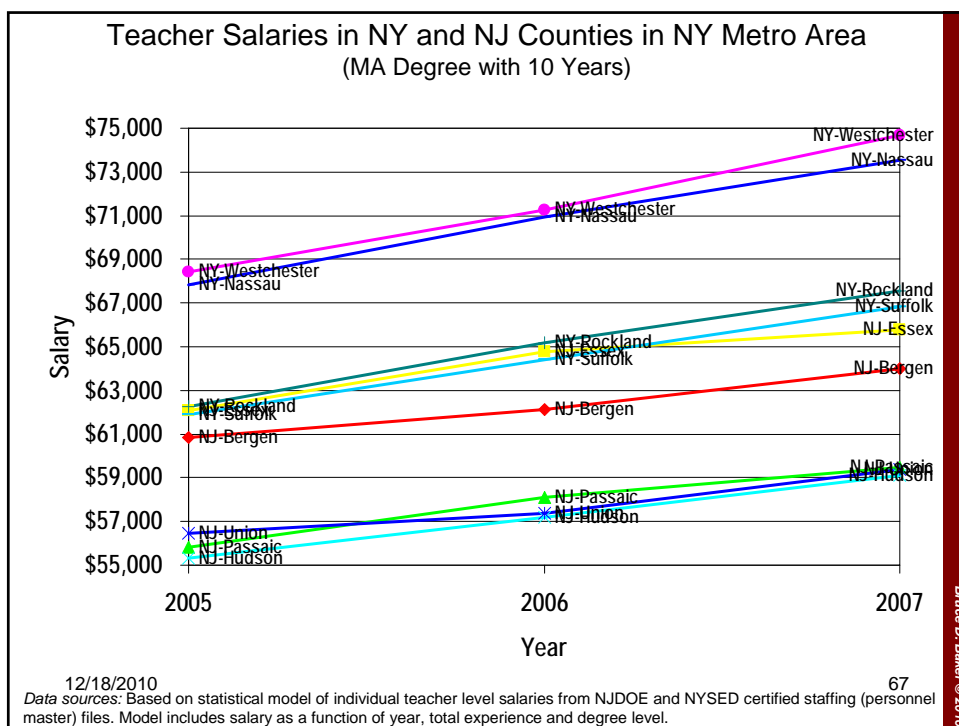
### Note regarding benefits & bias

- Corcoran and Mishel point out here: [http://epi.3cdn.net/05447667bb274f359e\\_zam6br3st.pdf](http://epi.3cdn.net/05447667bb274f359e_zam6br3st.pdf) that
  - “...overall K-12 teacher compensation was 27.5% greater than teacher wages alone, while overall professional compensation was 23.5% greater than professional wages. These differences in benefit shares translate into a benefits “bias” of 2.8 percentage points in 2006.”
- That is, benefits would close little of the overall gap in wages, even if the bias is somewhat larger in NJ.
- [Costrell and Podgursky](#) show about a 5% (slightly less) differential (10% non-teachers, 15% teachers) in the value of pensions, a portion of benefits. This too would close only part of the teacher to non-teacher wage gap in New Jersey, even if we assume New Jersey benefits for teachers to be much greater than other employee benefits.

12/18/2010

66

Bruce D. Baker © 2010



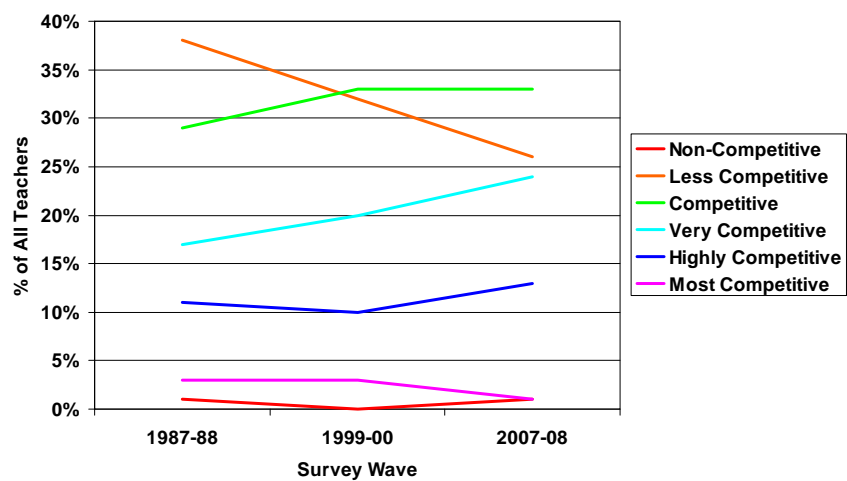
# Has the Quality in the Teacher Pipeline Changed in NJ?

12/18/2010

69

Bruce D. Baker © 2010

## Barron's Ratings of Undergraduate Institutions of New Jersey Teachers

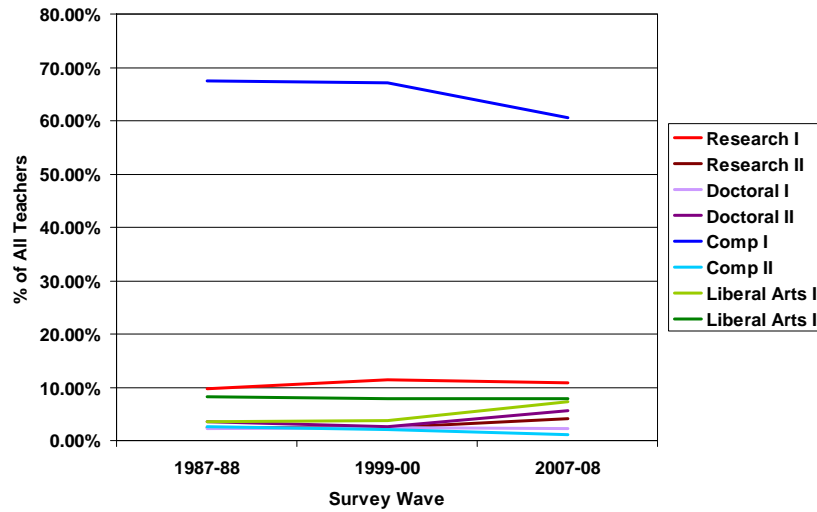


12/18/2010  
Data Source: NCES Schools and Staffing Survey

70

Bruce D. Baker © 2010

## Carnegie ('94) Classification of Undergraduate Institutions of New Jersey Teachers

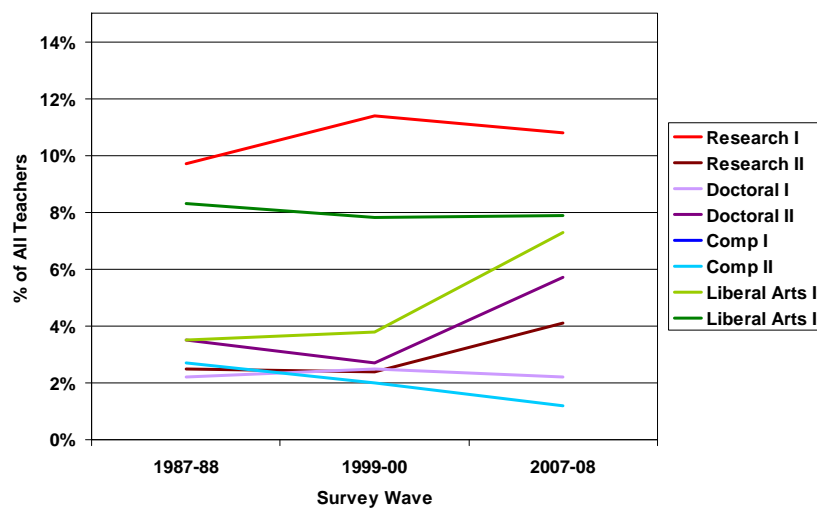


12/18/2010  
Data Source: NCES Schools and Staffing Survey

71

Bruce D. Baker © 2010

## Carnegie ('94) Classification of Undergraduate Institutions of New Jersey Teachers



12/18/2010  
Data Source: NCES Schools and Staffing Survey

72

Bruce D. Baker © 2010

## Enter the Jasons (Felch and Song) & the Los Angeles Times Value-Added Story



Bruce D. Baker © 2010

## Frequently used, factually incorrect statements about VAM

- ...a statistical approach known as value-added analysis, which rates teachers based on their students' progress on standardized tests from year to year. Each student's performance is compared with his or her own in past years, which largely controls for outside influences often blamed for academic failure: poverty, prior learning and other factors. (LA Times)
- VA measures “level the playing field for teachers who are assigned students of different ability.” (Kevin Carey, [here](#))
- “Value-added analysis can protect teachers from favoritism by using hard numbers and allow those with unorthodox methods to prove their worth.” (Kevin Carey, [here](#))

Bruce D. Baker © 2010

## Buddin's LAT Model

- Factors in the model
  - Prior year (not fall/spring) score
  - Student qualifies for free/red lunch (1=yes)
  - Student is limited English profic. (1 = yes)
  - Student joined school after Kindergarten.
  - Student gender
  - Year of data/test
  - Grade level of test
- Not included
  - **Composition of peer group**
  - **Multiple prior “lagged scores”**
  - **Disability status**
  - **Racial composition of class**
    - **Rockoff: many distrs dont control for race in value-added but b/c of achv gap this "makes it harder" for tchrs of AfAm students**
  - **Number of kids in class**
  - **A whole lot of other stuff**

12/18/2010

75

BRUCE D. BAKER © 2010

## Some fun findings from the LA Times

- 97% of children in the lowest performing schools are poor, and 55% in higher performing schools are poor;
- The number of gifted children a teacher has affects their value-added estimate, positively – The more gifted children the teacher has, the higher the effectiveness rating;
- Black teachers have lower value-added scores for both ELA and MATH than white teachers, and these are some of the largest negative correlates with effectiveness ratings provided in the report – especially for MATH.
- Having more black students in your class is negatively associated with teacher's value-added scores, though this effect is relatively small;
- Asian teachers have higher value-added scores than white teachers for Math, with the positive association between being Asian and math teaching effectiveness being as strong as the negative association for black teachers.

76

BRUCE D. BAKER © 2010

## Great Contradictions from the Jasons

- When asked whether “scale” issues - ceiling effects - influenced their analysis
  - *the Jasons replied that their finding that teachers with more gifted children had higher average “effectiveness” ratings provided evidence that ceiling effects weren’t a problem.*
- When asked whether value-added modeling could really control for the fact that kids aren’t randomly assigned across teachers
  - *The Jasons emphatically (though selectively) pointed toward the research of Kane and Staiger as providing an indisputable “yes!”*
- Wait... don’t these two statements contradict?

12/18/2010

77

Bruce D. Baker © 2010

## Notes for employment lawyers...

12/18/2010

78

Bruce D. Baker © 2010

# The next wave of lawsuits over teacher dismissal

Assume a policy/legislation is adopted permitting or **requiring removal** of tenure as a function of low “effectiveness” rating generated by value-added modeling...

12/18/2010

79

Bruce D. Baker © 2010

## Due Process Concerns

- To what extent are teachers provided sufficient information on how their ratings work and how/whether they can truly influence those ratings?
  - Major issue with DC IMPACT teacher guidebook
- To what extent might random error alone lead to teacher dismissal?
  - 10% to 20% chance of high performer being fired
  - 25% to 35% chance of average performer being fired
- To what extent might non-random assignment of students - totally outside the teacher's control - lead to dismissal?
- To what extent might more nefarious practices - like assigning tough kids to one teacher to increase chance of firing - lead to actual dismissal?

12/18/2010

80

Bruce D. Baker © 2010



## Title VII - Disparate Impact Claims

- Because of non-random assignment, and what we know about race/poverty, peer group effects, and the distribution of teachers by race with students by race, there may be strong patterns of racially disparate impact when dismissing teachers by VAM ratings.
  - In other words - black teachers are much more likely to be teaching poor black students, and therefore more likely to get lower VA ratings - hence be dismissed/de-tenured.
  - The crude - albeit typical - LAT model displays these differences.

12/18/2010

81

Bruce D. Baker © 2010

## Remedies/Alternatives

- Contractual protections for teachers
  - Random assignment clause
    - Stratified random assignment of all students to teachers, overseen by independent auditor
      - By race, gender, disability (by classific.), language, poverty, neighborhood, parent education, household chars. Etc.
  - Comparable conditions/resources clause
    - Room size/lighting/temp/location
    - Class meeting time of day (same and/or randomized)
    - Class size
- Less “discriminatory” alternatives
  - Basing VAM-related layoffs on within-race comparisons, and or within school (worst in group) norms for highly segregated schools
  - Including individual race and peer group race in VAM
  - Randomly assigning teachers by race across all schools and districts
  - Randomly assigning students by race across all teachers (schools and districts)

12/18/2010

82

Bruce D. Baker © 2010

## 3 Intriguing Policy Simulations

- Quality based RIF(rel. to seniority based)
  - Boyd, Lankford, Loeb and Wyckoff (Calder brief, 2010)
  - Assumes that if we lay off X teachers according to their value-added scores rather than seniority, we would see measurable improvements
  - <http://www.urban.org/UploadedPDF/1001421-teacher-layoffs.pdf>
- High Tenure Bar
  - Jonah Rockoff & Douglas Staiger
  - Assumes that if we only tenure the very top value-added producing teachers, holding all other factors constant, we could see significant improvements
  - <http://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.24.3.97>
- Dismissing the Bottom 5% to 10%
  - Eric Hanushek
  - Assumes that if we layoff systematically the bottom 5 to 10% teachers over time, we could see significant improvements.
  - [http://www.hks.harvard.edu/pepg/MeritPayPapers/Hanushek\\_10-13.pdf](http://www.hks.harvard.edu/pepg/MeritPayPapers/Hanushek_10-13.pdf)

12/18/2010

83

Bruce D. Baker © 2010

## Intriguing, yes, but... Assumptions in the Simulations

- All three of these simulations assume that the average quality of replacement/ applicants will necessarily be greater than those removed from the system.
  - And further, that if risk of dismissal is dramatically increased, and selectively increased in higher poverty, higher minority schools, that these policies will not make it even harder for these schools to recruit replacements.
- All three of these studies accept the possibility/ likelihood that a relatively large share of teachers will be misclassified as “failing” and dismissed, and that an equal share will be misclassified as “successful” and be retained. (collateral damage)
- None attempt to account for competitive replacement salary to improve average applicant quality (not an issue in RIF study)

12/18/2010

84

Bruce D. Baker © 2010

# Polis-Davis Title I Bill

- (B) REQUIREMENTS FOR TEACHER EVALUATIONS- The evaluation of a teacher's performance shall comply with the following minimum requirements:
- (i) A predominant factor in the evaluation is student academic growth with respect to the State's academic standards, as measured by–
- (I) student learning gains on the State's academic assessments established under paragraph (3) or another assessment of student academic achievement, as long as the assessment is the same for all students in the local educational agency in which the teacher is employed; and
- (II) if available, value-added measures that track individual student academic growth while under the instruction of the teacher.