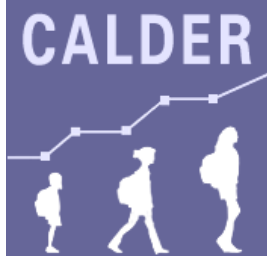


The Stability of Value-Added Measures of Teacher Quality and Implications for Teacher Compensation Policy

Tim R. Sass

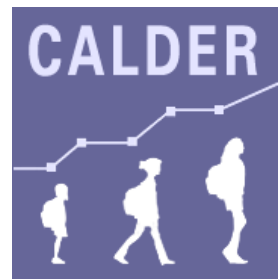
Florida State University and CALDER

November 21, 2008



Introduction

- Value-Added Measures of Teacher Quality
 - Measure impact of teacher on student achievement, holding constant other factors that affect student performance
 - Current and lagged student/family inputs, peer influences, school quality
 - Prior teachers
- For Value-Added Measures to Provide Correct Incentives They Must:
 - Be relatively accurate measures of productivity (ie. unbiased)
 - Be relatively stable over time



Introduction

□ Why Stability Matters for Policy

■ Tenure decisions

- If measured effects vary over time, could fail to retain many “good” teachers and keep many teachers who turn out to be relatively ineffective

■ Performance pay

- If value-added measures are unstable over time, compensation will be viewed as largely random and thus have little incentive effect



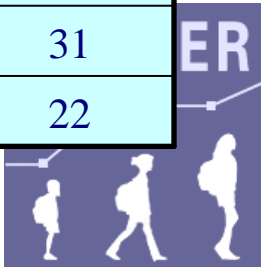
How Stable are Estimated Teacher Effects?

- ❑ Proportion of Top-Quintile Teachers Remaining in the Top Quintile the Next Year
- ❑ Year-to-Year Correlations

Quintile Rankings of Estimated Math Teacher Effects in 2000/01 and 2001/2002: Percent of Teachers by Row

		Ranking in 2001/02				
Ranking in 2000/01		Bottom 20%	Second 20%	Third 20%	Fourth 20%	Top 20%
Bottom 20%	San Diego, CA	35	25	16	14	11
	Duval Co., FL	30	20	20	12	18
	Hillsborough Co., FL	29	23	20	17	11
	Orange Co., FL	34	23	23	10	10
	Palm Beach Co., FL	24	12	22	26	16
Top 20%	San Diego, CA	12	9	25	24	29
	Duval Co., FL	14	13	22	25	27
	Hillsborough Co., FL	10	13	18	29	31
	Orange Co., FL	7	19	17	26	31
	Palm Beach Co., FL	13	18	18	20	22

Sources: Koedel and Betts (2007), McCaffrey, et al. (2008)



Year-to-Year Correlations in Estimated Teacher-by-Year Effects for Four Florida Counties

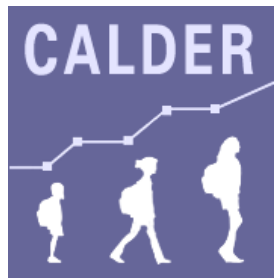
County	2000/01 & 2001/02	2001/02 & 2002/03	2002/03 & 2003/04	2003/04 & 2004/05
	Elementary			
Duval Co., FL	0.24	0.28	0.25	0.27
Hillsborough Co., FL	0.27	0.25	0.16	0.29
Orange Co., FL	0.31	0.34	0.30	0.36
Palm Beach Co., FL	0.16	0.08	0.21	0.21
	Middle			
Duval Co., FL	0.22	0.35	0.31	0.26
Hillsborough Co., FL	0.38	0.31	0.28	0.18
Orange Co., FL	0.36	0.28	0.31	0.24
Palm Beach Co., FL	0.24	0.32	0.28	0.26

Source: McCaffrey, et al. (2008)

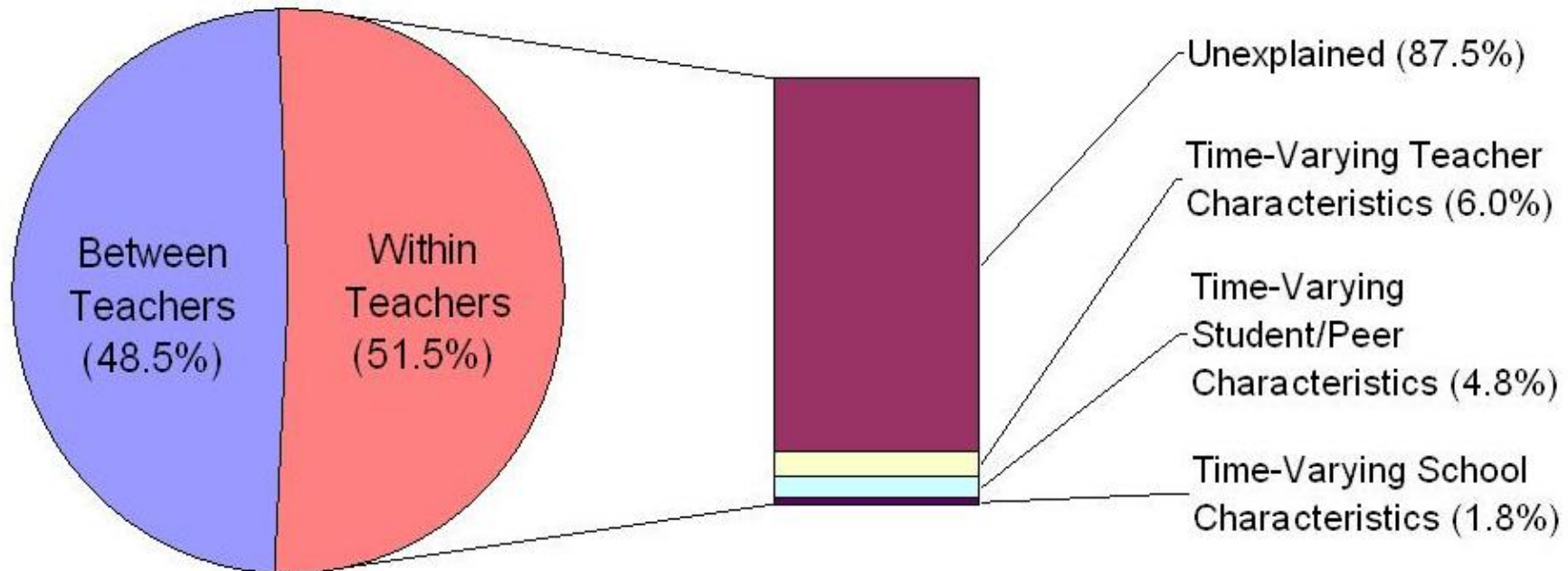


What Determines the Stability of Teacher Effect Estimates?

- ❑ Much of the Variation in Teacher Effects Over Time is Due to Student-Level Variation in Test Scores
 - Unexplained within-student variability in achievement gains
 - ❑ Random error or “noise”
 - “Dis-attenuated” correlations of teacher effects (which attempt to correct for student-level errors) much higher
 - ❑ 0.5 to 0.8 vs. 0.2 to 0.3 for raw correlations
 - Relatively small portion of within-teacher variation can be explained by observed time-varying student, peer, teacher and school characteristics
 - ❑ 12.5% in elementary school
 - ❑ 30.1% in middle school



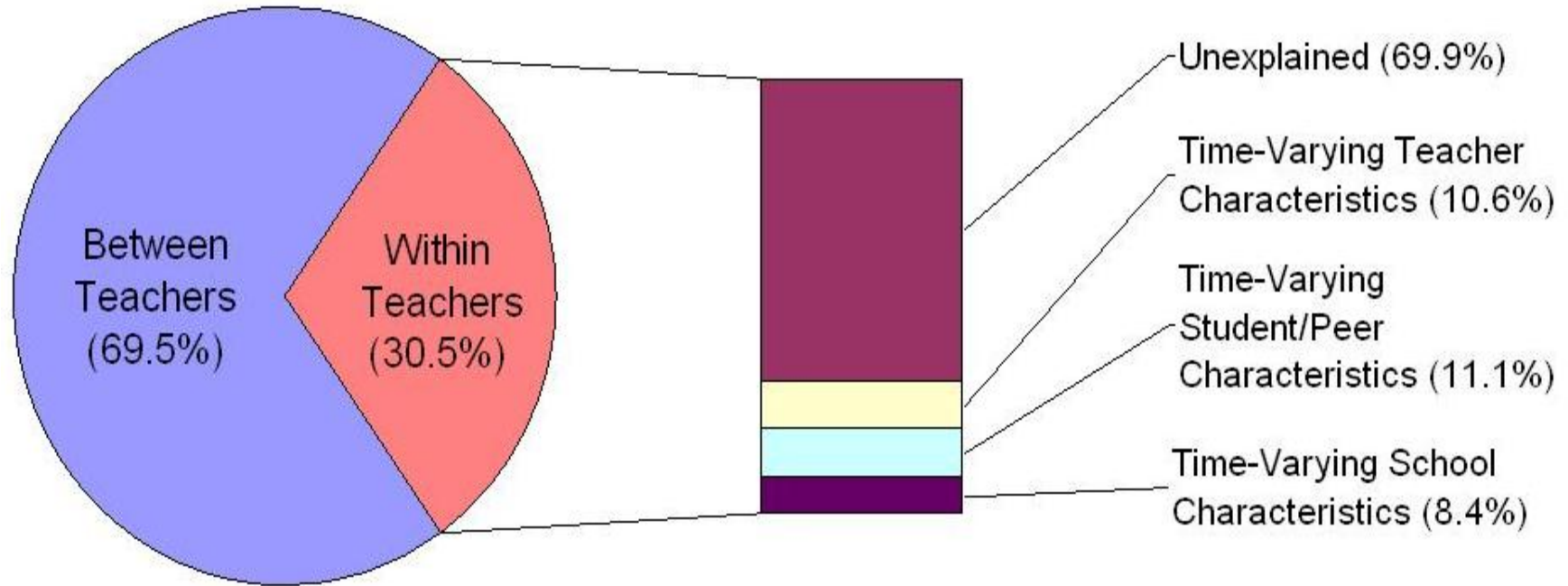
Components of Estimated Elementary Math Teacher-by-Year Effects: Simple Average of Duval, Hillsborough, Orange and Palm Beach Counties



Source: McCaffrey, et al. (2008)



Components of Estimated Middle School Math Teacher-by-Year Effects: Simple Average of Duval, Hillsborough, Orange and Palm Beach Counties



Source: McCaffrey, et al. (2008)



What Factors Affect the Stability of Estimated Teacher Effects?

□ Changes in the Test Instrument

■ Florida administers two tests in both math and reading

□ FCAT-SSS

- Criterion-reference exam
- “High-stakes” test used for school accountability

□ FCAT-NRT

- Norm-referenced exam (SAT-9/SAT-10)
- Essentially a “no-stakes” test in Florida

■ Teacher effect rankings vary by test, but variation is less than inter-temporal instability

- Could be due to differences in content, differential “test ceilings” or different responses to accountability pressures



Cross-Exam Stability of Elementary Math Teacher Effect Estimates by Quintile Percent of Teachers by Row) – Hillsborough Co., FL, 2001/02 [Correlation = 0.48]

	Ranking Based on SSS				
Ranking Based on NRT	Bottom 20%	Second 20%	Third 20%	Fourth 20%	Top 20%
Bottom 20%	43	26	14	11	5
Second 20%	26	25	23	17	10
Third 20%	15	21	21	24	18
Fourth 20%	11	19	25	20	24
Top 20%	4	9	17	27	43



Implications for Policy and Future Research

- ❑ Extent of Year-to-Year Variability in Estimated Teacher Effects Makes Use of Annual Value-Added Measures in Pay-for-Performance Systems Problematic
 - Hard to get stakeholders to buy into system that produces rankings that change radically from year to year
 - Likely little incentive effect if compensation is viewed as largely random and not closely tied to teacher effort

Implications for Policy and Future Research

□ Possible Alternatives

- Abandon attempts to have output-based incentives for teachers
 - Even noisy measures may be better than historical compensation schemes based on teacher inputs (experience and advanced degrees)
 - Evaluation of current pay-for performance systems will shed light on efficacy of outcome-based teacher compensation
 - Tennessee teacher incentive pay experiment
 - Teacher merit pay systems being implemented in Texas
 - Merit Award programs in Florida



Implications for Policy and Future Research

□ Possible Alternatives

■ Try to improve value-added measures

- Employ statistical methods to reduce the impact of student-level “noise” on estimates of teacher value added
 - measures may not be sufficiently transparent to stakeholders
- Use multi-year averages of teacher effects, rather than single-year value-added estimates
 - masks true year-to-year variation in teacher performance



Implications for Policy and Future Research

□ Possible Alternatives

- Combine value-added with principals' subjective evaluations of teachers
 - Principals may be able to observe many aspects of teacher performance that are not picked up by (noisy) test scores
 - Research by Jacob and Lefgren (2008) and by Harris and Sass (2007) finds principals can identify teachers who have the greatest impact on student achievement
 - Also find that principal evaluations are better at predicting future student achievement than are traditional measures used for teacher compensation (experience and educational attainment of teachers)
 - Determining accuracy and stability of teacher ratings from a hybrid system is an area for future research

CALDER

