

# Incentives, Selection, and Teacher Performance

Thomas S. Dee (Stanford GSE & NBER)

Jim Wyckoff (University of Virginia)

North Carolina Educator Effectiveness and  
Compensation Task Force

March 31, 2014

# Motivation

- A growing research consensus...
  - Large educational and economic returns to students when assigned to higher rather than lower-quality teachers
    - For example, replacing a bottom 5% teacher with an average teacher for 1 year increases lifetime income by an amount equal to \$250k in present value per classroom (Chetty et al. 2011)
  - But most easily observed teacher traits (e.g., experience, credentials) predict teacher quality only modestly.

# Motivation

- How to drive systematic improvements in teacher quality?
  - Interest in rethinking conventional teacher training and broadening pathways into the teaching profession
  - Substantial expenditures on teacher professional-development but little rigorous evidence on efficacy (Yoon et al. 2007)
  - Interest in teacher-compensation reforms
    - Teacher pay is currently based on rigid “single-salary” schedules that reward experience & credentials
    - Assessments of teacher performance are typically low-stakes “drive-by” without meaningful implications for pay, promotion, or targeted professional development

# The Diffusion of New Reforms

- Pilots that reconceive teacher evaluation, pay, and tenure
  - Strong federal encouragement through Race to the Top (RttT), Teacher Incentive Fund (TIF), School Improvement Grants (SIG), and NCLB Waivers
- A fitful and uncertain implementation?
  - Three big-city TIF awardees (NYC, Chicago, and Milwaukee) recently returned their funds because they could not secure teacher buy-in
  - Slow implementation of teacher evaluation in schools with SIG awards?
  - US DoED now allowing states with first NLCB waivers to delay teacher evaluation systems until AY 2016-17
  - State reforms complicated by the simultaneous roll-out of Common Core State Standards?

# Recent Teacher Incentive Experiments

- Project on Incentives in Teaching (POINT): 3-year RCT with middle-school math teachers in Nashville (Springer et al. 2010)
  - Bonuses of \$5k/10k/15k for average student performance at the 80<sup>th</sup>, 90<sup>th</sup>, and 95<sup>th</sup> percentile
  - No effects on student performance, teacher effort, classroom practice
- Group-based incentives of up to \$6k per teacher in a suburban TX district; no effects on student outcomes, teacher attitudes or practices (Springer et al. 2012)
- NYC (Fryer 2013): School-level RCT with rewards up to \$3k for meeting performance targets
  - No positive (possibly negative) effects on student outcomes
  - Most treatment schools chose group-based incentives
- Chicago Heights: group and individual incentives in 9 schools of as much as \$8k had no effects unless framed as a loss (Fryer et al. 2012)
- Teacher Advancement Program (TAP): random assignment to timing of implementation in 34 Chicago schools (Glazerman and Seifullah 2012)
  - Rewards up to \$6.4k for student achievement at the school or school-grade level and performance on classroom-observation rubric. Extra pay for increased responsibility
  - No effects on achievement but serious implementation problems

# Why Might Incentives Have No Effect?

- Teachers already highly motivated; marginal effect of incentives is slight?
- Test score targets were too ambitious to motivate teachers or weakly powered because they were often group-based?
- Incentives are muted when linked to a singular outcome (i.e., test scores) over which teachers don't perceive that they have direct control?
- Extant teachers lack the capacity or willingness to respond to incentives?
  - Better teacher supports are necessary and/or...
  - Differential recruitment of higher-quality teachers is more relevant than improving existing teachers.

# New Insights from DC Public Schools?

- IMPACT is the seminal, high-profile and uniquely ambitious performance assessment system introduced in DC Public Schools (DCPS) under then-Chancellor Michelle Rhee during AY 2009-10
- Teachers were evaluated annually on multiple measures and received an overall score [100, 400]



IMPACT thresholds 2009-10 to 2011-12

# Why is IMPACT Unique?

- *High-powered* and *individual* incentives (i.e., both large financial incentives but also *dismissal threats*)
- Teacher performance based on multiple measures (mostly classroom observation), not test scores alone (MET 2013)
- Implementation details: communication of expectations, rater training & data systems, teacher supports to meet expectations (e.g., instructional coaches)
- An at-scale, enduring system in a major urban school district, not a small-scale, temporary pilot



# IMPACT Timeline

- Fall 2006: Adrian Fenty elected Mayor of District of Columbia
  - January 2007: Fenty introduces legislation to restructure DCPS on first day in office
  - June 2007: Hours after officially assuming control of DCPS, Fenty announces Michelle Rhee as Chancellor
- March 2009: Rhee informs Council that a new teacher-evaluation system that includes student achievement data and classroom observation will be implemented in fall 2009
- Summer 2010: First DCPS teachers rated as Ineffective are dismissed; ME teachers under dismissal threat?
- September 2010: Fenty loses Democratic primary to Vincent Gray; Rhee resigns.
- March 2011: Kaya Henderson named Chancellor of DCPS
- Summer 2011: Teachers rated as Ineffective in 2010-11 are dismissed. *Teachers with second consecutive ME rating are dismissed for the first time.*

# Components of IMPACT Scores

- Teaching and Learning Framework (TLF): Classroom observations (3 by administrator, 2 by “master educator”) on rubric for effective instruction (e.g., well-organized, objective-driven lessons, checking for student understanding, etc.)
- Individual Value Added (IVA): *Growth* of teacher’s students on DC Comprehensive Assessment System (CAS) tests conditional on student and peer traits
- Teacher-Assessed Student Achievement (TAS): teachers choose (and administrators approve) learning goals based on non-CAS assessment(s)
  - At end of year, administrator rates success with rubric that privileges years of student learning or content mastery
- Commitment to School Community (CSC): administrator rates teacher support of school initiatives, promotion of high expectations, partnership with colleagues & parents
- School Value-Added Student Achievement (SVA): growth measure based on DC CAS

# Weighting of IMPACT Components\*

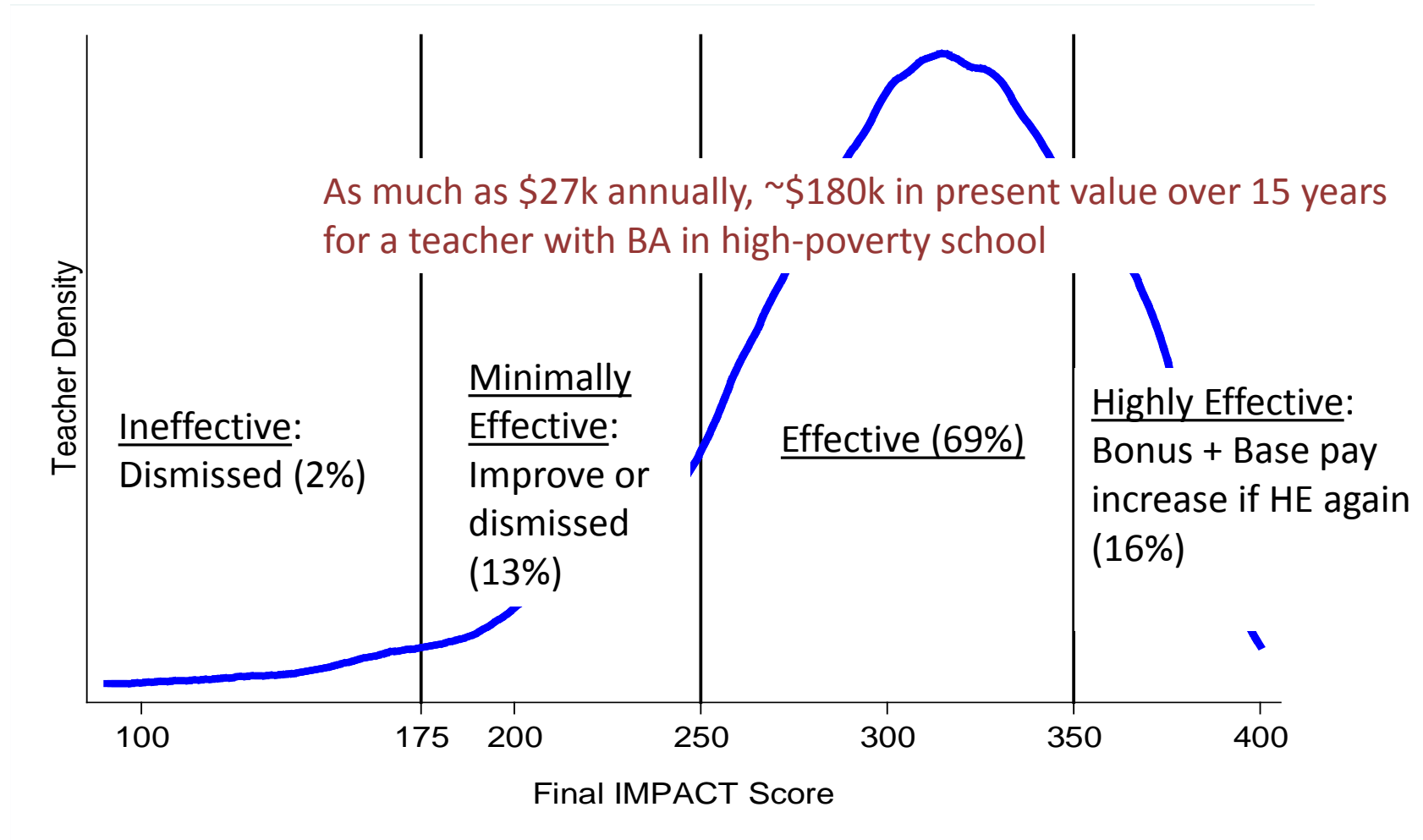
Component	Group 1**	Group 2
Individual Value Added	50%	0%
Teacher Learning Framework	35%	75%
Teacher Assessed Student Learning	0%	10%
Commitment to School Community	10%	10%
School Value Added	5%	5%

\* 2008-09 to 2011-12, the years of our analysis. Weights changed for 2012-13.

\*\*Fewer than 20 percent of DCPS teachers

- Core Professionalism (CP): administrator deductions (10-40 points) for poor attendance, punctuality, respect, adherence to policies and procedures

# How IMPACT Works\*



\*2008-09 to 2011-12, IMPACT rating groups changed for 2012-13

# Effects of IMPACT

IMPACT can improve performance of teachers through several mechanisms:

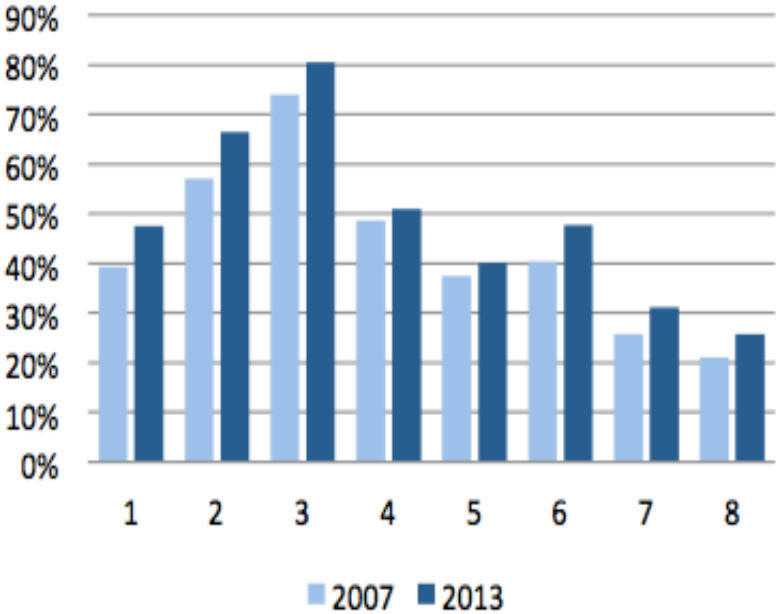
- Mechanical change in composition of teachers
  - teachers dismissed
- Behavioral responses to incentives
  - Voluntary exit of teachers
    - ME teachers increased attrition
    - HE teachers decreased attrition
  - Improvement in teacher performance
    - Both ME and HE teachers improve

# 2013 DC CAS Fast Facts

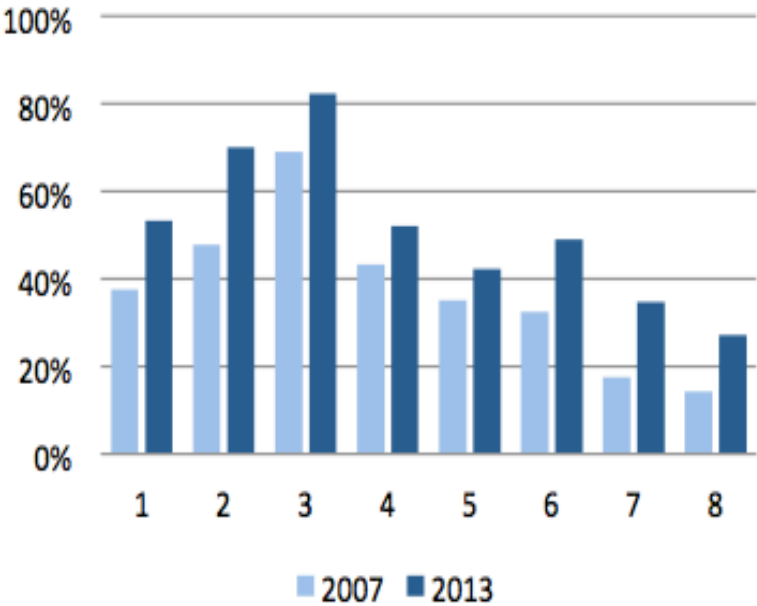
DCPS-ODS

Student in all 8 wards have improved proficiency rates in reading and math since 2007.

Reading Proficiency by Ward

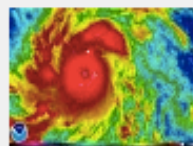


Math Proficiency by Ward



# Education

In the News Iran Obamacare Typhoon Haiyan Jobs report Xbox One 'Thor'



PHOTOS | The power of Haiyan



CBS admits major Benghazi gaffe



PH  
hai

## D.C. posts significant gains on national test, outpacing nearly every state



109



More ▾

By Emma Brown, Published: November 7 [E-mail the writer](#) ↵

The District's fourth- and eighth-graders made significant gains on national math and reading tests this year, posting increases that were among the city's largest in the history of the exam.

Students at the city's public charter and traditional schools also showed more improvement than nearly every state on the [2013 National Assessment of Educational Progress](#), an exam the federal government has administered every two years since the early 1990s. Results of the tests, known as the Nation's Report Card, were released Thursday morning.

# Teachers meaningfully respond to IMPACT incentives

Teachers	Attrition	IMPACT Score
Minimally effective under threat of dismissal	+50%	+12.6
Highly Effective with possibility of base-pay increase	+, not significant	+10.9

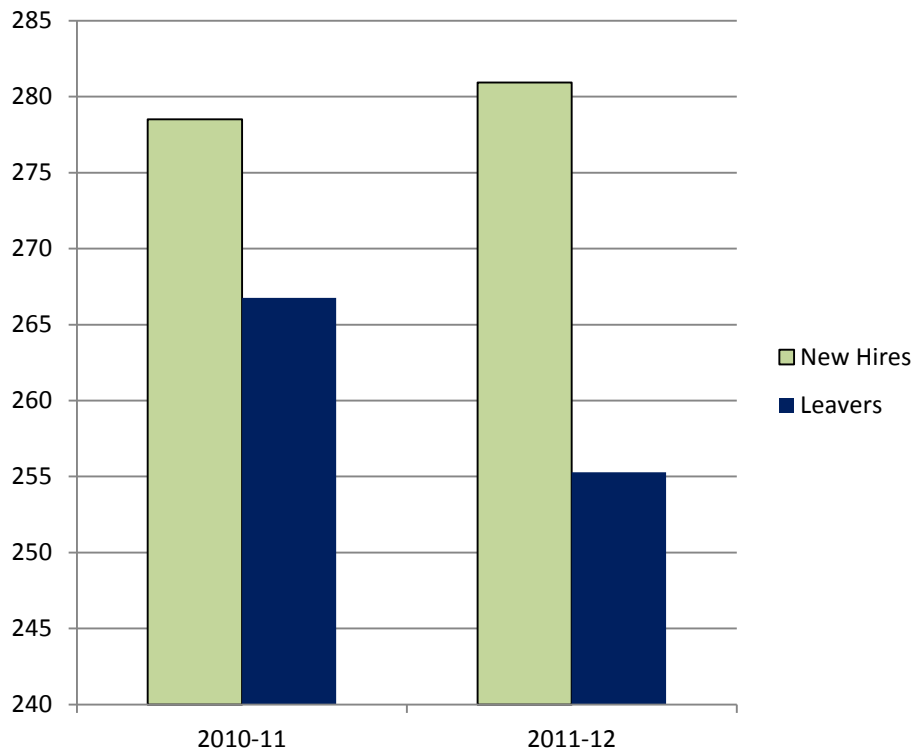
- ME teachers have attrition that is 11 percentage points higher than otherwise identical teachers not facing dismissal threat (=50 percent increase)
- ME teachers who remain improve by 12.6 IMPACT points; equal to about half the gain novice teachers make in 3 years
- Once Highly Effective teachers improve by 10.9 IMPACT points; equal to 40 percent of gain novice teachers make in 3 years



# Alternative explanations?

- Causal warrant of the RD design generally appears quite strong (e.g., density tests, covariate balance, bandwidth restrictions, “placebo” RD)
- Non-random teacher mobility to different school, to non-teaching position with IMPACT scores, to a different teaching “group”?
  - Auxiliary RD regressions indicate that ME/HE had statistically insignificant effects on group status in  $t+1$  and mobility from teaching position in original school
- Differential cheating on DC CAS by teachers?
  - Allegations of broader test manipulation actually precede IMPACT
  - Omitting small number of IMPACT-era teachers identified by test-security firm doesn’t change results.
  - Performance gains exist for non-test IMPACT components

# IMPACT Scores of Entering and Leaving Teachers, 2010-11 & 2011-12



New hires substantially outperformed all teachers who left DCPS, suggesting the composition of the DCPS workforce meaningfully improved

# Discussion

- Evidence that a performance-assessment system with high-powered incentives meaningfully influenced teacher retention and performance
- Interestingly, IMPACT performance of *new hires* clearly exceeds that of leavers (supply of teachers?)
- Currently analyzing effects on student achievement

# Possible lessons for other states & districts

- A compelling *proof point* for this type of human-capital strategy?
- Important lessons on effective design & implementation (a cautionary tale?)
  - Target multiple measures, not just test scores
  - Clear communication of expectations/measurement
  - Careful rater training
  - Complementary data systems
  - Policy credibility and leadership buy-in