The Effects of Performance Pay Bonuses for Top Teachers on Student Test Scores

Josh Hollinger

2/6/20

Motivation

Motivation

- If teacher quality matters, how can teacher quality be increased, particularly in "high-need" schools?

Motivation

- If teacher quality matters, how can teacher quality be increased, particularly in "high-need" schools?
- Can performance pay (rewarding value-added) incentivize teachers to improve?

Motivation

- If teacher quality matters, how can teacher quality be increased, particularly in “high-need” schools?
- Can performance pay (rewarding value-added) incentivize teachers to improve?
- The literature in the US thus far suggests maybe not. (Springer et al. 2010, Fryer 2013)

Motivation

- One way forward: identify design flaws empirically
  - Evidence of reasons for weak incentives
  - Evidence of response where incentives should be stronger (Imberman and Lovenheim 2015)

Motivation

- One way forward: identify design flaws empirically
  - Evidence of reasons for weak incentives
  - Evidence of response where incentives should be stronger (Imberman and Lovenheim 2015)
- Common feature: performance threshold
  - Could be too low or too high
  - Generally set high
  - Strongest incentive for "marginal" teachers

Research Question

Do teachers predicted to have a better chance of attaining a performance bonus improve their students' test scores more?

Overview of This Paper

- Evaluates effect of teacher performance pay on student test scores in 21 “high-need” elementary schools in North Carolina

Overview of This Paper

- Evaluates effect of teacher performance pay on student test scores in 21 "high-need" elementary schools in North Carolina
- Predicts probability of teachers attaining the value-added required to receive bonus

Overview of This Paper

- Evaluates effect of teacher performance pay on student test scores in 21 "high-need" elementary schools in North Carolina
- Predicts probability of teachers attaining the value-added required to receive bonus
- Investigates if teachers with higher probability of bonus respond more to incentive

Preview of Results

- No evidence of overall effect: perhaps positive math, negative reading

Preview of Results

- No evidence of overall effect: perhaps positive math, negative reading
- No evidence of better results for teachers with higher probability of bonus (some contrary)

Preview of Results

- No evidence of overall effect: perhaps positive math, negative reading
- No evidence of better results for teachers with higher probability of bonus (some contrary)
  - Single-year VA estimates are quite noisy
  - Do teachers have any idea how changing their effort changes their VA?
  - Are teachers already motivated?
  - Should we focus more on the bottom of the distribution?

Literature

- Effects of performance pay in education

Literature

- Effects of performance pay in education
  - Group-based: Often found positive effects in international settings (Lavy 2002; Glewwe et al. 2003) and no effects in the US (Fryer 2013; Goodman and Turner 2013)

Literature

- Effects of performance pay in education
  - Group-based: Often found positive effects in international settings (Lavy 2002; Glewwe et al. 2003) and no effects in the US (Fryer 2013; Goodman and Turner 2013)
  - Individual: some evidence of positive effects for a few programs (Dee and Wyckoff 2015, Eren 2019), but experimental designs have found no effects (Fryer 2012, Springer et al. 2010)

Literature

- Effects of performance pay in education
  - Group-based: Often found positive effects in international settings (Lavy 2002; Glewwe et al. 2003) and no effects in the US (Fryer 2013; Goodman and Turner 2013)
  - Individual: some evidence of positive effects for a few programs (Dee and Wyckoff 2015, Eren 2019), but experimental designs have found no effects (Fryer 2012, Springer et al. 2010)
  - Springer et al. 2010: were high thresholds a problem?

Literature

- Effects of performance pay in education
  - Group-based: Often found positive effects in international settings (Lavy 2002; Glewwe et al. 2003) and no effects in the US (Fryer 2013; Goodman and Turner 2013)
  - Individual: some evidence of positive effects for a few programs (Dee and Wyckoff 2015, Eren 2019), but experimental designs have found no effects (Fryer 2012, Springer et al. 2010)
  - Springer et al. 2010: were high thresholds a problem?
  - Hill and Jones (2018): no effects for high schools, but male teachers improved

Literature

- Effects of performance pay in education
  - Group-based: Often found positive effects in international settings (Lavy 2002; Glewwe et al. 2003) and no effects in the US (Fryer 2013; Goodman and Turner 2013)
  - Individual: some evidence of positive effects for a few programs (Dee and Wyckoff 2015, Eren 2019), but experimental designs have found no effects (Fryer 2012, Springer et al. 2010)
  - Springer et al. 2010: were high thresholds a problem?
  - Hill and Jones (2018): no effects for high schools, but male teachers improved
- Non-linear incentives

Literature

- Effects of performance pay in education
  - Group-based: Often found positive effects in international settings (Lavy 2002; Glewwe et al. 2003) and no effects in the US (Fryer 2013; Goodman and Turner 2013)
  - Individual: some evidence of positive effects for a few programs (Dee and Wyckoff 2015, Eren 2019), but experimental designs have found no effects (Fryer 2012, Springer et al. 2010)
  - Springer et al. 2010: were high thresholds a problem?
  - Hill and Jones (2018): no effects for high schools, but male teachers improved
- Non-linear incentives
  - Sales commissions and executive compensation: Oyer (1998), Larkin (2014)

Literature

- Effects of performance pay in education
  - Group-based: Often found positive effects in international settings (Lavy 2002; Glewwe et al. 2003) and no effects in the US (Fryer 2013; Goodman and Turner 2013)
  - Individual: some evidence of positive effects for a few programs (Dee and Wyckoff 2015, Eren 2019), but experimental designs have found no effects (Fryer 2012, Springer et al. 2010)
  - Springer et al. 2010: were high thresholds a problem?
  - Hill and Jones (2018): no effects for high schools, but male teachers improved
- Non-linear incentives
  - Sales commissions and executive compensation: Oyer (1998), Larkin (2014)
  - Brehm et al. (2017): closest to my paper

Outline of This Paper

- Theoretical Framework
- Performance Pay Programs
- Data
- Policy Evaluation
- Predicting Bonus Probability
- Effects by Bonus Probability
- Robustness
- Conclusion

Theoretical Framework - adapted from Neal (2011)

- Principal-agent model: employ teacher to teach a student
- Output is test score y, teacher effort t unobserved:

$$y = t + \epsilon$$

- ϵ distributed N(0,σ^2)
- Without performance pay, teacher chooses $t = \bar{t}$
- Quadratic cost of additional effort:

$$f(t) = c(t - \bar{t})^2$$

Theoretical Framework

- Offer bonus b if test score exceeds a threshold y^T
- Wage is $w = s + b * 1(y > y^T)$
- Teacher's utility is $U = w - c(t - \bar{t})^2$
- Teacher maximizes with t:

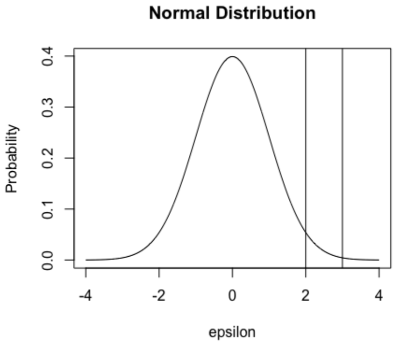
$$b[1 - \Phi(y^T - t)] - c(t - \bar{t})^2$$

First order condition:

$$b\phi(y^T - t) = c(t - \bar{t})$$

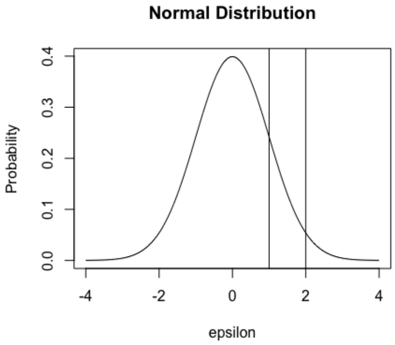
Theoretical Framework

What happens if target is set too high?



Theoretical Framework

More within reach: higher marginal return to effort



Performance Pay Policies

- Teacher Incentive Fund gave money to districts to implement teacher incentives in high-need schools

Performance Pay Policies

- Teacher Incentive Fund gave money to districts to implement teacher incentives in high-need schools
- Two large districts in NC (Charlotte-Mecklenburg and Guilford) received grants to implement performance pay in a subset of their schools deemed high-need

Performance Pay Policies

- Teacher Incentive Fund gave money to districts to implement teacher incentives in high-need schools
- Two large districts in NC (Charlotte-Mecklenburg and Guilford) received grants to implement performance pay in a subset of their schools deemed high-need
- Charlotte-Mecklenburg's VA bonuses started in 2010, Guilford in 2007

Performance Pay Policies

- Teacher Incentive Fund gave money to districts to implement teacher incentives in high-need schools
- Two large districts in NC (Charlotte-Mecklenburg and Guilford) received grants to implement performance pay in a subset of their schools deemed high-need
- Charlotte-Mecklenburg's VA bonuses started in 2010, Guilford in 2007
- Bonuses based on individual teacher's VA in a given year: had to exceed a VA threshold set relative to all teachers of the same grade in the district

Performance Pay Policies

- Teacher Incentive Fund gave money to districts to implement teacher incentives in high-need schools
- Two large districts in NC (Charlotte-Mecklenburg and Guilford) received grants to implement performance pay in a subset of their schools deemed high-need
- Charlotte-Mecklenburg's VA bonuses started in 2010, Guilford in 2007
- Bonuses based on individual teacher's VA in a given year: had to exceed a VA threshold set relative to all teachers of the same grade in the district
  - Charlotte-Mecklenburg: $2500 if above 70th percentile in the district
  - Guilford: $2000 if 1 SD above district mean, $6000 if 1.5 SD above

Other Incentives Implemented

- Recruitment incentives
- Guilford: teachers with VA 2 SD below the mean 2 years in a row were moved to a different school in the district
- Principals could also earn performance bonuses

Data

- North Carolina Education Research Data Center
- All 3rd-5th grade students in NC 2000-2011
- Students' end-of-grade test scores, demographic characteristics
- Identifiers allow tracking students and teachers over time and linking students to teachers and schools
- Student-teacher link defined by the teacher administering test
  - Student's classroom teacher unless absent
  - Keep only teachers known to be teaching math/reading in the right grade

Value-Added: general concept

Estimate fixed-effect for each teacher:

$$y_{icsjt} = \theta y_{ics,t-1} + \beta X_{it} + \gamma Z_{ct} + \delta S_{st} + \mu_j + \epsilon_{icsjt}$$

i=student, c=class, s=school, j=teacher, t=year

Value-Added: as implemented

- Estimated separately for each year
- Multilevel hierarchical linear model (mixed model)
- Random coefficients for teacher and school
- Two lags of math and reading

$$y_{icsjt} = \theta y_{ics,t-1} + \beta X_{it} + \gamma Z_{ct} + \delta S_{st} + \mu_j + \mu_s + \epsilon_{icsjt}$$

i=student, c=class, s=school, j=teacher, t=year

Value-added

- I use the list of controls used in the policies and a mixed model to match how VA is calculated for bonuses
- Highly correlated with more basic FE model (0.905)
- Don't observe who actually gets bonuses, but estimate from VA and threshold rules
- Estimate about 17% of eligible teachers get bonus

Summary Statistics

Summary Statistics		
VARIABLES	**North Carolina** mean (SD)	**Performance Pay Schools** mean (SD)
Student Characteristics		
Math Score	0.057	-0.459
	(0.983)	(0.959)
Reading Score	0.051	-0.510
	(0.981)	(0.984)
White	0.603	0.114
	(0.489)	(0.318)
Black	0.263	0.725
	(0.440)	(0.446)
Hispanic	0.067	0.081
	(0.251)	(0.273)
Female	0.495	0.503
	(0.500)	(0.500)
Economically Disadvantaged	0.451	0.822
	(0.498)	(0.383)
Exceptional	0.114	0.142
	(0.317)	(0.349)
Limited English	0.0248	0.0414
	(0.156)	(0.199)
Teacher Characteristics		
Experience	12.51	9.198
	(9.587)	(9.067)
White	0.865	0.602
	(0.341)	(0.490)
Black	0.119	0.385
	(0.324)	(0.487)
Female	0.925	0.902
	(0.263)	(0.297)
VA	-0.00298	-0.0238
	(1.016)	(1.061)
Observations	1,633,497	16,035

Empirical Framework

- Difference-in-differences: student is "treated" if their school has performance pay in that year, "ever treated" if their school has performance pay at some time in the panel

Empirical Framework

- Difference-in-differences: student is "treated" if their school has performance pay in that year, "ever treated" if their school has performance pay at some time in the panel

$$y_{isgt} = \theta treated_{st} + \Gamma ever\_treated_s + \beta X_{it} + \delta_{gt} + \epsilon_{isgt}$$

Empirical Framework

- Difference-in-differences: student is "treated" if their school has performance pay in that year, "ever treated" if their school has performance pay at some time in the panel

$$y_{isgt} = \theta treated_{st} + \Gamma ever\_treated_{s} + \beta X_{it} + \delta_{gt} + \epsilon_{isgt}$$

- Identifying assumption: no year- and treatment-group-specific shocks besides performance pay

Performance Pay Effects - Math

Effect of Performance Pay Programs on Math Scores

	(1)	(2)	(3)	(4)
Treated	0.0536*	0.000125	0.0523	0.0437
	(0.0280)	(0.0306)	(0.0325)	(0.0386)
Observations	1,658,694	1,658,507	1,658,693	1,658,425
R-squared	0.695	0.735	0.702	0.738
Student controls	YES	YES	YES	YES
Year-by-grade FE	YES	YES	YES	YES
Teacher FE		YES		
Teacher-by-School FE				YES
School FE			YES	

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Standard errors are clustered at the school level.

Performance Pay Effects - Reading

Effect of Performance Pay on Reading Scores

	(1)	(2)	(3)	(4)
Treated	-0.0130	-0.0751***	-0.0119	-0.0644*
	(0.0126)	(0.0274)	(0.0147)	(0.0369)
Observations	1,658,694	1,658,507	1,658,693	1,658,425
R-squared	0.669	0.689	0.673	0.692
Student controls	YES	YES	YES	YES
Year-by-grade FE	YES	YES	YES	YES
Teacher FE		YES		
Teacher-by-School FE				YES
School FE			YES	

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Standard errors are clustered at the school level.

Performance Pay Effects - Math - Class controls

Effect of Performance Pay on Math Scores

	(1)	(2)	(3)	(4)
Treated	0.0646**	0.00493	0.0613*	0.0474
	(0.0290)	(0.0310)	(0.0338)	(0.0389)
Observations	1,658,694	1,658,507	1,658,693	1,658,425
R-squared	0.696	0.735	0.703	0.738
Student controls	YES	YES	YES	YES
Class controls	YES	YES	YES	YES
Year-by-grade FE	YES	YES	YES	YES
Teacher FE		YES		
Teacher-by-School FE				YES
School FE			YES	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Standard errors are clustered at the school level.

Performance Pay Effects - Reading - Class controls

Effect of Performance Pay on Reading Scores

	(1)	(2)	(3)	(4)
Treated	-0.00313	-0.0733***	-0.00778	-0.0638*
	(0.0123)	(0.0272)	(0.0145)	(0.0363)
Observations	1,658,694	1,658,507	1,658,693	1,658,425
R-squared	0.670	0.689	0.673	0.692
Student controls	YES	YES	YES	YES
Class controls	YES	YES	YES	YES
Year-by-grade FE	YES	YES	YES	YES
Teacher FE		YES		
Teacher-by-School FE				YES
School FE			YES	

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Standard errors are clustered at the school level.

Predicting Probability of Bonus

Predicting Probability of Bonus

- Predict each teacher's probability of attaining a bonus (1 SD above their district mean VA)

Predicting Probability of Bonus

- Predict each teacher's probability of attaining a bonus (1 SD above their district mean VA)
- Treated teacher-years are excluded from regressions, then their VA is predicted by coefficients on rest of sample

Predicting Probability of Bonus

- Predict each teacher's probability of attaining a bonus (1 SD above their district mean VA)
- Treated teacher-years are excluded from regressions, then their VA is predicted by coefficients on rest of sample
- What do teachers know?
    - Past VA (and if above 1 SD in the past)
    - Class characteristics
    - Teacher characteristics
    - All of the above
    - Average test scores in the past
    - Average test score gains
    - What if they think bonus is based on test score levels or gains?

Predicting Probability of Bonus

Linear Probability Model using Past VA:

$$VA_{jt} = \sum_{y=1}^{5} [\beta_{1y} 1(VA_{j,t-y} > 1SD) + \beta_{2y} VA_{j,t-y} + \beta_{3y} VA^2_{j,t-y}] + \epsilon_{jt}$$

Predicting Probability of Bonus

Linear Probability Model using Past VA:

$$VA_{jt} = \sum_{y=1}^{5} [\beta_{1y} 1(VA_{j,t-y} > 1SD) + \beta_{2y} VA_{j,t-y} + \beta_{3y} VA_{j,t-y}^2] + \epsilon_{jt}$$

- Include indicators for missing lags
- Can add class and teacher characteristics, C_{jt} and T_{jt}
- Also, instead of VA, can use lags of average test scores / gains to predict VA

Predicting Probability of Bonus

Linear Probability Model using Past VA:

$$VA_{jt} = \sum_{y=1}^{5} [\beta_{1y} 1(VA_{j,t-y} > 1SD) + \beta_{2y} VA_{j,t-y} + \beta_{3y} VA^2_{j,t-y}] + \epsilon_{jt}$$

- Include indicators for missing lags
- Can add class and teacher characteristics, C_{jt} and T_{jt}
- Also, instead of VA, can use lags of average test scores / gains to predict VA Lastly, can predict probability of getting bonus if it were defined by test score levels / gains

Predicting Probability of Bonus - Regression Results

Predicting Probability of Bonus

Past VA variables	1(Bonus)	Teacher	1(Bonus)	Class	1(Bonus)
Lag Bonus	0.0681***	No experience	-0.0694***	Class size	-0.00113***
	(0.00721)		(0.00538)		(0.000229)
Lag VA	0.0672***	1-3 years exp.	-0.0140***	Male	-0.0263***
	(0.00212)		(0.00419)		(0.00881)
Lag VA^2	0.0145***	4-9 years exp.	-0.0121***	Age	0.00196
	(0.00122)		(0.00370)		(0.00578)
Lag 2 Bonus	0.0673***	10-24 year exp.	-0.00220	Limited English	-0.0236
	(0.00769)		(0.00348)		(0.0153)
Lag 2 VA	0.0339***	>25 years exp.	-	Exceptional	0.00445
	(0.00225)				(0.00807)
Lag 2 VA^2	0.00899***	Licensing exam	0.00751***	Disadvantaged	-0.00584
	(0.00135)		(0.00160)		(0.00528)
Lag 3 Bonus	0.0347***	Certified	0.00316	Repeat	-0.00533
	(0.00885)		(0.00409)		(0.0189)
Lag 3 VA	0.0158***	Male	0.00161	1st year	-0.0228***
	(0.00271)		(0.00401)		(0.00575)
Lag 3 VA^2	0.00441***	Black	-0.0561**	Lag Math	-0.0199***
	(0.00161)		(0.0251)		(0.00425)
Lag 4 Bonus	0.0395***	Hispanic	-0.0638*	Lag Reading	0.0105**
	(0.00995)		(0.0327)		(0.00453)
Lag 4 VA	0.0150***	Am. Indian	-0.0302	Lag Absences	-0.000653
	(0.00328)		(0.0286)		(0.000466)
Lag 4 VA^2	0.00282	Multi-race	-0.0880***	Lag Days ISS	-0.0332**
	(0.00184)		(0.0311)		(0.0147)
Lag 5 Bonus	0.0241**	White	-0.0471*	Lag Days OSS	-0.0130*
	(0.0113)		(0.0249)		(0.00743)
Lag 5 VA	0.0145***			Lag Times ISS	0.0580
	(0.00366)				(0.0357)
Lag 5 VA^2	-0.000448			Lag Times OSS	0.0500**
	(0.00205)				(0.0222)
				AIG math	0.0256*
					(0.0154)
				AIG reading	-0.0368**
					(0.0159)
Constant	0.184***	Observations	102,395	R-squared	0.099

R-squared for Different Models

R-squared of Predictions	
Prediction Model	
Full Model	9.85%
Past VA and Teacher variables	9.76%
Past VA variables	9.56%
Past Gains variables	7.37%
Past Levels variables	1.51%
Class variables	0.09%
Levels predicting levels	15.26%

Effects by Probability of Bonus

Testing for heterogeneous effects by bonus probability:

$$y_{ijsgt} = \theta_1 Tr_{st} + \theta_2 Tr_{st} \times Prob_{jt} + \Gamma_1 EvTr_s + \Gamma_2 EvTr_s \times Prob_{jt} + \nu Prob_{jt} + \beta X_{it} + \delta_{gt} + \epsilon_{isgt}$$

Effects by Probability of Bonus

Effect on Test Scores by Predicted Probability of Bonus

Specifications	Math	Reading
Linear		
Treated $\times$ Pr(Bonus)	-0.027	-0.273**
	(0.144)	(0.109)
Quadratic		
Treated $\times$ Pr(Bonus)	0.060	-0.302
	(0.211)	(0.205)
Treated $\times$ Pr(Bonus)^2	-0.514	-0.010
	(0.681)	(0.685)
Above/Below Median		
Treated $\times$ [Pr(Bonus) $<$ 13.6%]	0.083**	0.043**
	(0.0417)	(0.0214)
Treated $\times$ [Pr(Bonus) $\geq$ 13.6%]	0.053**	-0.041**
	(0.0313)	(0.0208)
Terciles		
Treated $\times$ [Pr(Bonus) $<$ 10.7%]	0.081*	0.050
	(0.0426)	(0.0319)
Treated $\times$ [10.7 % $\leq$ Pr(Bonus) $<$ 17.2%]	0.089**	0.006
	(0.0420)	(0.0136)
Treated $\times$ [Pr(Bonus) $\geq$ 17.2%]	0.089**	-0.025
	(0.0345)	(0.0271)
Observations	1,658,694	1,658,694
Student controls	YES	YES
Year-by-grade FE	YES	YES

Using Test Score Gains

Effect on Test Scores by Predicted Probability of "Gains' Bonus

Specifications	Math	Reading
Linear		
Treated × Pr(Bonus)	-0.078	-0.303*
	(0.224)	(0.159)
Quadratic		
Treated × Pr(Bonus)	-0.062	-0.240
	(0.398)	(0.309)
Treated × Pr(Bonus)^2	-0.090	-0.304
	(1.483)	(0.913)
Above/Below Median		
Treated × [Pr(Bonus) $<$ 13.4%]	0.076**	0.021
	(0.0350)	(0.0147)
Treated × [Pr(Bonus) $\geq$ 13.4%]	0.088***	-0.007
	(0.0324)	(0.0226)
Terciles		
Treated × [Pr(Bonus) $<$ 11.5%]	0.097**	0.040
	(0.0468)	(0.0316)
Treated × [11.5 % $\leq$ Pr(Bonus) $<$ 15.9%]	0.101***	0.037*
	(0.0377)	(0.0219)
Treated × [Pr(Bonus) $\geq$ 15.9%]	0.032	-0.045*
	(0.0351)	(0.0268)
Observations	1,658,694	1,658,694
Student controls	YES	YES
Year-by-grade FE	YES	YES

Using Test Score Levels

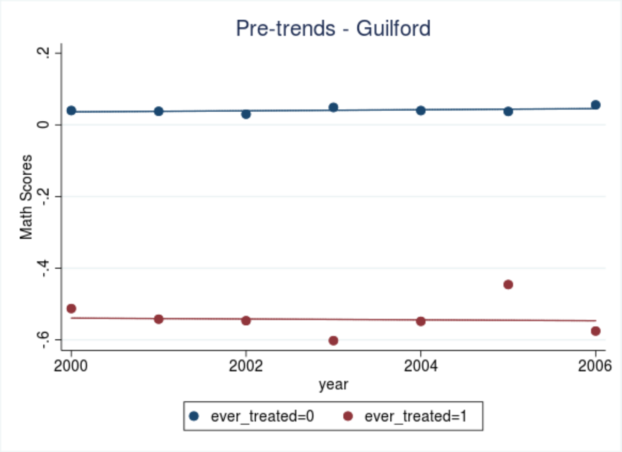
Effect on Test Scores by Predicted Probability of "Levels" Bonus

Specifications	Math	Reading
Linear		
Treated $\times$ Pr(Bonus)	-0.329	-0.536**
	(0.342)	(0.222)
Quadratic		
Treated $\times$ Pr(Bonus)	0.142	-0.382
	(0.259)	(0.253)
Treated $\times$ Pr(Bonus)^2	2.397*	0.652
	((1.239))	(1.322)
Above/Below Median		
Treated $\times$ [Pr(Bonus) $<$ 12.9%]	0.041	-0.006
	(0.0327)	(0.0139)
Treated $\times$ [Pr(Bonus) $\geq$ 12.9%]	0.145***	-0.013
	(0.0429)	(0.0543)
Terciles		
Treated $\times$ [Pr(Bonus) $<$ 9.8%]	0.027	-0.004
	(0.0338)	(0.0141)
Treated $\times$ [9.8 % $\leq$ Pr(Bonus) $<$ 14.3%]	0.106***	0.026
	(0.0316)	(0.0304)
Treated $\times$ [Pr(Bonus) $\geq$ 14.3%]	0.014	-0.098**
	(0.0533)	(0.0436)
Observations	1,658,694	1,658,694
Student controls	YES	YES
Year-by-grade FE	YES	YES

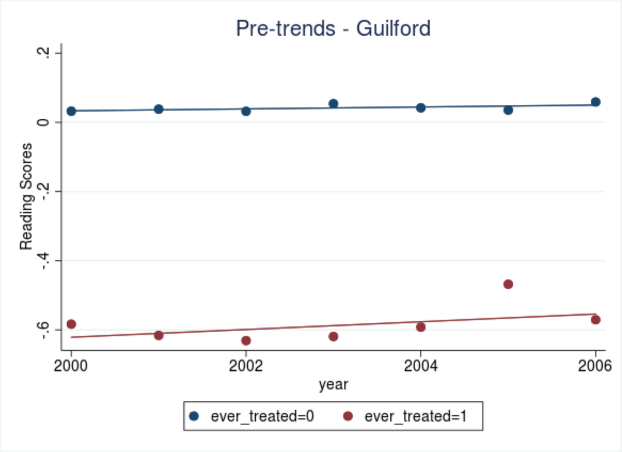
Robustness

- Pre-trends: separately for each district
- Differential classroom composition in treatment group
- Differential sorting of students into classrooms

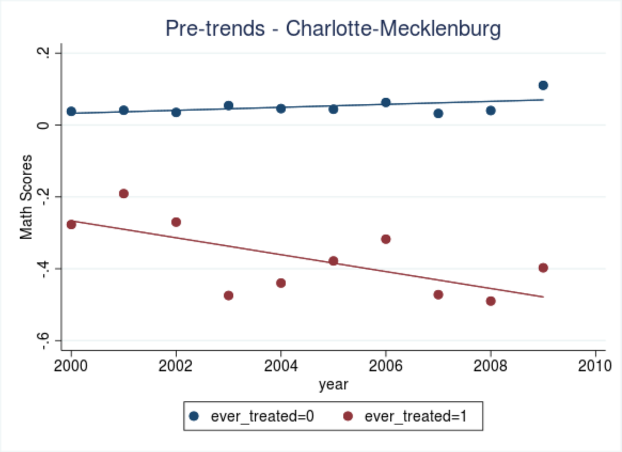
Pre-Trends: Guilford Math



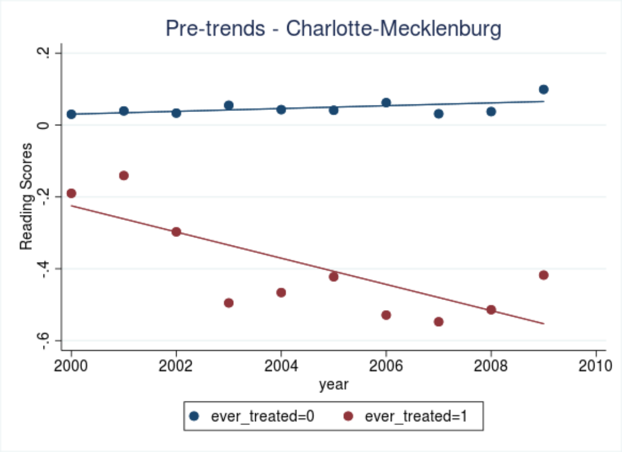
Pre-Trends: Guilford Reading



Pre-Trends: Charlotte-Mecklenburg Math



Pre-Trends: Charlotte-Mecklenburg Math



Pre-Trend Regression

Testing for Pre-trends in Math Scores

VARIABLES	(1)	(2)	(3)	(4)
Ever-treated × Time	0.00560	0.00430	0.0153	0.00883
	(0.00458)	(0.00614)	(0.0106)	(0.0118)
Observations	1,099,634	1,099,462	1,099,630	1,099,406
R-squared	0.696	0.738	0.705	0.740
Student controls	YES	YES	YES	YES
Year-by-grade FE	YES	YES	YES	YES
Teacher FE		YES		
Teacher-by-School FE				YES
School FE			YES	

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Standard errors are clustered at the school level.

Pre-Trend Regression

Testing for Pre-trends in Reading Scores

VARIABLES	(1)	(2)	(3)	(4)
Ever-treated × Time	-0.00410	-0.00289	0.0111	0.00130
	(0.00409)	(0.00566)	(0.00847)	(0.0106)
Observations	1,099,634	1,099,462	1,099,630	1,099,406
R-squared	0.696	0.738	0.705	0.740
Student controls	YES	YES	YES	YES
Year-by-grade FE	YES	YES	YES	YES
Teacher FE		YES		
Teacher-by-School FE				YES
School FE			YES	

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Standard errors are clustered at the school level.

Classroom Characteristics

Classroom Characteristics and Treatment

Characteristic	"Treatment Effect"
Age	-0.092***
	(0.0202)
AIG math	-0.041**
	(0.0166)
AIG reading	-0.040**
	(0.0161)
Exceptional	-0.036***
	(0.0125)
First year in school	0.0349
	(0.0232)
Lag Days Absent	0.762***
	(0.249)
Lag Days ISS	0.006
	(0.015)
Lag Days OSS	0.095***
	(0.0365)
Lag Times ISS	0.008
	(0.0142)
Lag Times OSS	0.033***
	(0.0125)
Lag Math	-0.011
	(0.0526)
Lag Reading	-0.094*
	(0.0535)
Limited English	0.000
	(0.00981)
Economically Disadvantaged	0.016
	(0.0183)
Male	-0.004
	(0.0139)
Repeat grade	-0.006***
	(0.00208)
Observations (Teacher-year)	102,664

Sorting Students into Classrooms

Testing for Sorting to Teachers within Schools

Characteristic	Treatment effect on SD
Age	-0.096
	(0.103)
AIG math	-0.013
	(0.0140)
AIG reading	-0.014
	(0.0130)
Exceptional	-0.036**
	(0.0176)
First year in school	0.079**
	(0.0365)
Lag Days Absent	6.360***
	(0.585)
Lag Days ISS	0.874***
	(0.0469)
Lag Days OSS	0.953***
	(0.0513)
Lag Times ISS	0.874***
	(0.0469)
Lag Times OSS	0.882***
	(0.0380)
Lag Math	0.028
	(0.0641)
Lag Reading	-0.080
	(0.0722)
Limited English	0.011
	(0.0147)
Economically Disadvantaged	0.052*
	(0.0266)
Male	-0.013
	(0.0271)
Repeat grade	-0.008*
	(0.00432)
Observations (School-year)	15,236

Conclusion

- Overall effect of performance pay seems small

Conclusion

- Overall effect of performance pay seems small
- I fail to find evidence for teachers closer to the threshold responding more to the incentive

Conclusion

- Overall effect of performance pay seems small
- I fail to find evidence for teachers closer to the threshold responding more to the incentive
- Why?
  - Teachers may not have any idea what their predicted VA is or how/if their effort changes their VA

Conclusion

- Overall effect of performance pay seems small
- I fail to find evidence for teachers closer to the threshold responding more to the incentive
- Why?
    - Teachers may not have any idea what their predicted VA is or how/if their effort changes their VA
    - The large noise in single-year VA may weaken incentives
        - Makes it hard to learn from one's VA
        - Makes it hard to get teachers to buy in

Conclusion

- Overall effect of performance pay seems small
- I fail to find evidence for teachers closer to the threshold responding more to the incentive
- Why?
  - Teachers may not have any idea what their predicted VA is or how/if their effort changes their VA
  - The large noise in single-year VA may weaken incentives
    - Makes it hard to learn from one's VA
    - Makes it hard to get teachers to buy in
  - Incentive targets high-performing teachers; do they have less margin for improvement?

Conclusion

- Overall effect of performance pay seems small
- I fail to find evidence for teachers closer to the threshold responding more to the incentive
- Why?
  - Teachers may not have any idea what their predicted VA is or how/if their effort changes their VA
  - The large noise in single-year VA may weaken incentives
    - Makes it hard to learn from one's VA
    - Makes it hard to get teachers to buy in
  - Incentive targets high-performing teachers; do they have less margin for improvement?
  - Are teachers already sufficiently focused on test scores? (NCLB)

Possible Solutions

- Use multiple years to evaluate teacher VA, incorporate into salary increases

Possible Solutions

- Use multiple years to evaluate teacher VA, incorporate into salary increases
- Evaluate small groups of teachers to encourage teamwork and reduce noise

Possible Solutions

- Use multiple years to evaluate teacher VA, incorporate into salary increases
- Evaluate small groups of teachers to encourage teamwork and reduce noise
- Focus more incentives on bottom of the teacher VA distribution