Economics Bulletin

Volume 36, Issue 2

The Effects of a Teacher Performance-Pay Program on Student Achievement: A Regression Discontinuity Approach

Yusuke Jinnai
International University of Japan

Abstract

This paper presents evidence from a regression-discontinuity analysis of a teacher performance-pay program, in which teachers are awarded an additional cash bonus for improving their students' achievement. Results show that teachers who failed to reach an expected benchmark for their students' achievement, resulting in no bonuses, performed significantly better in the subsequent year than those who reached this benchmark and thus received a bonus. This finding highlights that the presence of performance-pay incentives affects student achievement in future years by inducing more effort from teachers who failed in the present year. Moreover, the results demonstrate that such impact disappeared once the government repealed the pay scheme: another indication that teachers actively respond to monetary bonuses.

This research was financially supported by The Seimeikai Foundation and JSPS KAKENHI (Grant Number 26780179). I am grateful to Greg Caetano, Josh Kinsler, and Ronni Pavan for their helpful comments. I also benefited from discussion with Tom Ahn, Jacob Vigdor, and Jeffrey Zabel. All remaining errors are my own.
Citation: Yusuke Jinnai, (2016) "The Effects of a Teacher Performance-Pay Program on Student Achievement: A Regression Discontinuity Approach", *Economics Bulletin*, Volume 36, Issue 2, pages 993-999
Contact: Yusuke Jinnai - yjinnai@iuj.ac.jp.
Submitted: January 28, 2016. **Published:** June 11, 2016.

1. Introduction

Coupled with school choice, school accountability has been a centerpiece of public education reform in the United States as well as other developed countries for the past decade. Accountability programs, such as the one used under the No Child Left Behind (NCLB) Act of 2001 in the U.S., are designed to provide useful information to parents and legislators so that they can effectively monitor the performance of each school. These programs typically evaluate schools based on student achievement on statewide standardized tests and assign simple ratings (e.g. A to F) for public reporting. With a variety of rewards for high-performing schools as well as sanctions for low-performing ones, accountability programs have increased student achievement as policymakers have anticipated (Hanushek and Raymond, 2005; Jacob, 2005; Figlio and Rouse, 2006; Reback, 2008).

In this study, I address the effects of the accountability program established in 1996 in the state of North Carolina, U.S. In contrast to NCLB and many other programs that set *level* targets, North Carolina's sophisticated system sets *growth* targets that take into account prior scores to adjust for students' diverse characteristics and family backgrounds. Moreover, the program provides each teacher with up to $1,500 per year as incentives for improving student achievement.

The main question in this study is how receiving incentive bonuses (or failing to receive such bonuses) affects teachers and schools in the subsequent year. To answer this question, I estimate the impact of receiving bonuses on student achievement by employing the regression discontinuity (RD) design with a threshold that separates bonus-qualified schools and non-qualified ones. Under the RD design, those schools around the threshold can be considered almost identical in school quality and other characteristics; however, only the teachers at schools which exceed the threshold receive bonuses. Therefore, the difference in the following year's student achievement between schools just above and just below the threshold can be attributed to whether or not teachers received cash bonuses in the present year.

In the analysis, I use detailed data sets from North Carolina and divide them into three stages. The first period consists of two school years 2005-06 and 2006-07, when qualified teachers were awarded the maximum bonus of $1,500. In 2007-08, the second period, North Carolina reduced the maximum amount to $1,053. Because of its continuing worse economic condition, in the end, the state repealed its bonus system in the 2008-09 school year, which is the third period of my study. By separating the sample into these three different stages, this study estimates the impact of the accountability program with (i) *full* incentives, (ii) *reduced* incentives, and (iii) *no* incentives, respectively. In particular, the period under no bonuses is used for a placebo test that estimates the impact without such monetary incentives.

Estimation results demonstrate that schools where teachers did not receive bonuses performed significantly better in the following year than schools where teachers did. Bonus non-qualified schools improved their average academic performance by 0.06-0.09 standard deviations compared to qualified schools. Moreover, the placebo test highlights that such impact disappeared once the state government repealed the pay scheme: another indication that teachers are responsive to cash bonuses.

This paper contributes to the literature of teacher incentive programs (such as Ahn and Vigdor (2014) and Vigdor (2008)) by providing two new empirical findings. First, this study shows that the positive impact of the incentive program on bonus non-qualified schools is persistent over time even after incentives are reduced. Second, the finding from the placebo test illustrates that teachers are sensitive to monetary rewards but not to (non-monetary) school ratings without such bonuses.

2. North Carolina's accountability program

North Carolina provides a particularly good setting for examining the effects of an accountability system because it has had a carefully designed educational accountability system in place since academic year 1996-97. Of particular significance, the North Carolina accountability program evaluates schools primarily on the annual achievement gains of their students from one year to the next. This growth approach to accountability aims at leveling the playing field for all students; for instance, students from economically disadvantaged and minority families tend to perform worse on tests than those from more affluent families. Because of its focus on individual growth, North Carolina's model is considered more sophisticated than level models, which judge schools on the average level of test scores.[1]

For each year, North Carolina's system first standardizes each student's test score and then calculates each student's gain as a standard deviation unit. Finally, for each school, the system computes the school-level average growth in test scores across all students in all subjects.[2] In elementary and middle schools, if a school's average growth score is equal to or greater than zero, the school is qualified for the school-wide bonuses of up to $1,500. In 2007-08, however, the maximum amount of bonus was reduced to $1,053 due to North Carolina's worse economic condition. Although teachers had taught their classes with the expectation of the full bonus of $1,500, they were notified of this reduction at the end of the academic year. Moreover, incentive bonuses have been suspended since 2008-09, as the state budget remains limited.

These exogenous policy changes provide a unique opportunity to examine the impact of teacher incentives on their performance. In particular, such policies allow this study to distinguish between the impact of a performance-pay program with different amount of incentives (i.e., full, reduced, or zero bonuses).

3. Data

3.1. Summary statistics

In this study I use data for school years from 2005-06 to 2009-10 because the new policies implemented in 2005-06 make comparisons to previous years inappropriate. Detailed data sets on students, teachers, and schools are provided by the North Carolina Education Research Data Center, and school-level average growth scores are provided by the North Carolina Department of Public Instruction. Table 1 shows the summary statistics for public schools in North Carolina in the 2005-06 school year. While 914 schools qualified for bonuses (i.e., average growth score ≥ 0), other 851 schools were non-qualified (i.e., average growth score < 0). As expected, there are also remarkable differences in academic achievement

[1] Ladd and Lauen (2010) argue that growth models also have an advantage that they are less vulnerable to gaming behaviors than level models.

[2] Details of the program are available at North Carolina's ABCs Accountability Model (http://abcs.ncpublicschools.org/abcs/).

captured by other measurements between bonus qualified and non-qualified schools. For instance, the proportions of schools that met Adequate Yearly Progress (AYP), the primary measurement used under the federal No Child Left Behind, is 0.58 for qualified schools and 0.31 for non-qualified ones. Although the differences in school characteristics are relatively small, those in student characteristics between the two kinds of schools are noticeable. Bonus qualified schools tend to have more white students, fewer black students, and fewer students who are eligible for free or reduced-price lunch programs. Regarding teacher characteristics, bonus qualified schools have more teachers with advanced degrees and face lower turnover rates.

3.2. Graphical evidence of RD design

Since this study is primarily interested in whether a school is qualified for bonuses (i.e. average growth score ≥ 0), I focus on the data around the threshold where the average growth score equals zero. First, Figure 1 shows an indicator for bonus receipt in year 2005-06, which equals one if a school received bonuses and zero otherwise. As expected, the figure demonstrates that all schools with average growth score above zero received bonuses ($= 1$), and those with average growth score below zero did not ($= 0$), resulting in a sharp RD design.

4. Identification

Following Imbens and Lemieux (2008), this study employs the following regression equation for a sharp RD design:

$$Score_{s,t+1} = \beta_0 + \beta_1 D_{st} + \beta_2 Score_{st} + \beta_3 D_{st} Score_{st} + X_{s,t+1}\gamma + \epsilon, \tag{1}$$

where $Score_{st}$ is the average growth score of school s in year t, and $D_{st} = 1$ (treatment) if school s receives bonuses in year t; otherwise $D_{st} = 0$ (control). Samples are limited to $-h < Score_{st} < h$ for a given bandwidth h. The interaction term $D_{st}Score_{st}$ allows the estimation equation to have a different slope for each side of the threshold. School characteristics are represented by $X_{s,t+1}$ in year $t+1$, and I show estimation results with and without these control variables.

The estimated coefficient β_1 captures the impact of teachers' receiving bonuses on outcome variables. Critical to the estimation under the RD design is the choice of bandwidth h, and this is informed by using Silverman's rule of thumb, which is suggested as an initial choice by Imbens and Kalyanaraman (2012). The optimal bandwidth is calculated as $h = h_0 = 1.06 * sd * N^{-1/5}$, where sd denotes the standard deviation of the treatment-defining variable and N denotes the number of observations. I also use a wider bandwidth $h_1 = \frac{3}{2}h_0$ as well as a shorter bandwidth $h_2 = \frac{2}{3}h_0$ to check whether the estimation results are sensitive to the choice of bandwidth.

5. Results

Table 2 demonstrates the results from RD estimation for the period under full bonus incentives ($1,500): the average treatment effect on 2006-07 and 2007-08 outcomes. Columns

(1) and (2) show the results with the widest bandwidth $h_1 = \frac{3}{2}h_0$, Columns (3) and (4) show those with the baseline bandwidth h_0 chosen by Silverman's rule, and Columns (5) and (6) show those with the shortest bandwidth $h_2 = \frac{2}{3}h_0$.

Column (1) shows the estimate of -0.0565 with significance at the 1% level, implying that schools that did not receive bonuses significantly increased their academic performance by 0.0565 standard deviations in the following year, compared to bonus-qualified schools. Column (2) shows the estimation result with control variables $X_{s,t+1}$ in equation (1), which include school characteristics such as class size and school enrollment, student characteristics such as the proportions of white and black students, and teacher characteristics such as the proportion of teachers with advanced degrees. The estimate is not significantly different from that in Column (1).

In a similar manner, Columns (3) and (4) with the baseline bandwidth h_0 as well as Columns (5) and (6) with the shorter bandwidth h_2 demonstrate that the estimates are significant and negative, suggesting that bonus non-qualified schools performed significantly better in the subsequent year than bonus qualified schools.

Under the reduced bonus incentives, the results show a similar pattern: significant and negative effects. This finding suggests that even after the reduction in the maximum amount of the bonuses (from \$1,500 to \$1,053), teachers at bonus non-qualified schools performed significantly better than those at qualified schools in the following year. Since the magnitude of the impact is similar between the full bonus incentives and the reduced bonus incentives, I use these three years to estimate the average impact of the bonus receipt, which is the primary result in this study.

Table 3 demonstrates the main results: the average treatment effect of bonus receipt. With the largest bandwidth $h_1 = 0.0507$, estimates are highly significant around -0.060 either with or without controls (Columns 1 and 2). With Silverman's optimal bandwidth $h_0 = 0.0338$, the results are also highly significant around -0.075 (Columns 3 and 4). The results with the shortest bandwidth $h_2 = 0.0225$ are less significant, but still significant and the effects are negative (Columns 5 and 6) as same as the previous results. Although the magnitude of the effects ranges from around -0.06 to -0.09, the impact of bonus receipt is significant, implying that bonus non-qualified schools achieved significantly higher academic performance than qualified schools in the subsequent year.

By contrast, however, the placebo test with no monetary incentives highlights that, once the state government repeals cash bonuses, there is no such effect that induces additional performance from non-qualified schools and teachers. Table 4 shows insignificant effects on 2009-10 outcome when incentive bonuses are repealed. Strikingly, under no bonus incentives, all of the estimates in the table are not significant for any bandwidth. This finding suggests that once monetary incentives are removed, teachers at low-achieving schools do not exert any more additional effort than those at high-achieving schools.

Another implication is that school ratings themselves do not induce more effort from teachers or schools. Even without such bonuses, schools receive a rating such as "Less than Expected" if they do not meet the threshold of zero average growth. However, the insignificant estimates in Table 4 underline that schools and teachers are not sensitive to school ratings when there are no monetary incentives. In sum, Table 3 and Table 4 shed light on the fact that low-performing schools expend additional effort only when they are provided with cash-bonus incentives.

6. Conclusion

While a large number of states and countries have introduced school accountability programs and teacher performance pay systems, recent studies have found that these policies do not always accomplish their expected results. In this paper, I examined the impact of teachers' receiving incentive bonuses on student achievement in the subsequent year. Despite the fact that all schools and teachers are subject to the same incentive scheme each year, this study finds that (1) bonus non-qualified schools performed significantly better than qualified schools in the following year, (2) the same pattern was also observed even under the reduced bonus incentives, and (3) such impact disappeared when the state government repealed the bonus program.

The effects found in this study can stem from gaming behaviors (Figlio, 2006), but Ladd and Lauen (2010) argue that growth models are less vulnerable to such behaviors than other evaluation models, and therefore teacher effort seems the likely explanation for the estimated impact. The new empirical findings demonstrated in this paper contribute to the growing literature on performance pay for public-school teachers.

References

Ahn, T., and J. Vigdor. 2014. "When incentives matter too much: Explaining significant responses to irrelevant information." *NBER Working Paper No. 20321*, 1–44.

Figlio, D. 2006. "Testing, crime and punishment." *Journal of Public Economics*, 90(4-5): 837–851.

Figlio, D., and C. Rouse. 2006. "Do accountability and voucher threats improve low-performing schools?" *Journal of Public Economics*, 90(1-2): 239–255.

Hanushek, E., and M. Raymond. 2005. "Does school accountability lead to improved student performance?" *Journal of Policy Analysis and Management*, 24(2): 297–327.

Imbens, G., and K. Kalyanaraman. 2012. "Optimal bandwidth choice for the regression discontinuity estimator." *Review of Economic Studies*, 79(3): 933–959.

Imbens, G., and T. Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics*, 142(2): 615–635.

Jacob, B. 2005. "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools." *Journal of Public Economics*, 89(5-6): 761–796.

Ladd, H., and D. Lauen. 2010. "Status versus growth: The distributional effects of school accountability policies." *Journal of Policy Analysis and Management*, 29(3): 426–450.

Reback, R. 2008. "Teaching to the rating: School accountability and the distribution of student achievement." *Journal of Public Economics*, 92(5-6): 1394–1415.

Vigdor, J. 2008. "Teacher salary bonuses in North Carolina." *Working Paper at National Center for Analysis of Longitudinal Data in Education Research.*

Figure 1: Assignment to bonus receipt

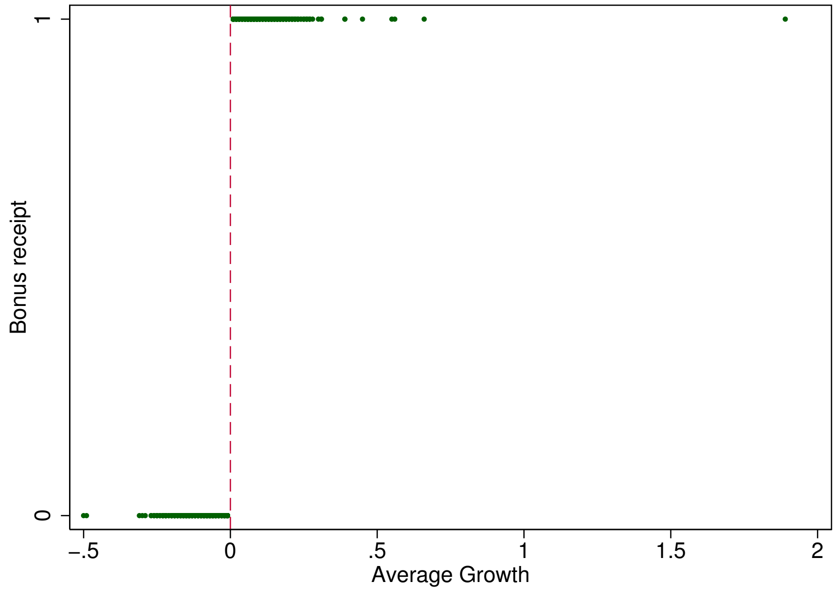

Table 1: Summary statistics for public schools in North Carolina in 2005-06

	Bonus qualified (average growth ≥ 0)		Bonus non-qualified (average growth < 0)	
	mean	s.d.	mean	s.d.
Academic achievement				
Average growth score	0.082	0.091	-0.082	0.062
AYP met(%)	0.58	0.49	0.31	0.46
School characteristics				
Enrollment	558.1	248.3	530.7	231.5
Class size	20.1	2.66	19.7	2.69
Student-teacher ratio	14.9	4.24	14.9	4.68
Student characteristics				
White(%)	0.617	0.262	0.467	0.290
Black(%)	0.264	0.227	0.400	0.271
Hispanic(%)	0.085	0.089	0.097	0.099
Free lunch eligible(%)	0.372	0.190	0.495	0.194
Teacher characteristics				
Advanced degree(%)	0.269	0.095	0.228	0.089
Turnover(%)	0.202	0.101	0.225	0.106
N	914		851	

Note: The figures do not include high schools or non-regular schools that follow different rules for school ratings and bonus receipt.

Table 2: RD estimates with full bonus incentives (impact on 2006-07 and 2007-08)

	bandwidth h_1		bandwidth h_0		bandwidth h_2	
	(1)	(2)	(3)	(4)	(5)	(6)
D_{st}	-0.0565***	-0.0513**	-0.0619**	-0.0524*	-0.0274**	-0.0249*
	(0.0209)	(0.0207)	(0.0303)	(0.0299)	(0.0134)	(0.0133)
Control $X_{s,t+1}$	No	Yes	No	Yes	No	Yes
N	858	822	558	535	286	274

Note: The dependent variable is school-level average growth. Standard errors clustered by school in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 3: RD estimates with bonus incentives (impact on 2006-07 through 2008-09)

	bandwidth h_1		bandwidth h_0		bandwidth h_2	
	(1)	(2)	(3)	(4)	(5)	(6)
D_{st}	-0.0612***	-0.0594***	-0.0790***	-0.0721***	-0.0956**	-0.0847*
	(0.0147)	(0.0145)	(0.0238)	(0.0236)	(0.0396)	(0.0437)
Control $X_{s,t+1}$	No	Yes	No	Yes	No	Yes
N	1238	1188	667	637	364	347

Note: The dependent variable is school-level average growth. Standard errors clustered by school in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 4: RD estimates with no bonus incentives (impact on 2009-10)

	bandwidth h_1		bandwidth h_0		bandwidth h_2	
	(1)	(2)	(3)	(4)	(5)	(6)
D_{st}	0.0176	0.00569	0.0168	0.0116	0.0368	0.0439
	(0.0230)	(0.0234)	(0.0319)	(0.0322)	(0.0466)	(0.0481)
Control $X_{s,t+1}$	No	Yes	No	Yes	No	Yes
N	305	288	203	191	137	129

Note: The dependent variable is school-level average growth. Standard errors in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.