Economics & Management Series EMS-2016-08

To Introduce or Not To Introduce Monetary Bonuses: The Cost of Repealing Teacher Incentives

Yusuke Jinnai
International University of Japan

January 2016

IUJ Research Institute
International University of Japan

To Introduce or Not To Introduce Monetary Bonuses: The Cost of Repealing Teacher Incentives*

Yusuke Jinnai†

January, 2016

Abstract

Teacher performance pay programs form the foundation of recent reforms in public education. Although existing research has found monetary bonuses for teachers increase student achievement, no studies have examined the potentially negative effects of repealing such incentives. Using novel data from North Carolina, where the state government first reduced and finally repealed its teacher incentive program, this paper shows that student achievement at the lowest-performing schools significantly decreased after the reduction in bonuses and further decreased after the repeal of the incentive program. These findings illustrate that once incentives are introduced it is not cost-free to reduce or remove them.

JEL classification: I21, H4

Keywords: School accountability, Performance pay, Teacher incentives

*I am grateful to Greg Caetano, Josh Kinsler, and Ronni Pavan for their guidance and extensive feedback. I also thank Kara Bonneau for her help in obtaining data from the North Carolina Education Research Data Center (NCERDC). Financial support from The Seimeikai Foundation and JSPS KAKENHI (Grant Number 26780179) is gratefully acknowledged. All errors are my own.

†Graduate School of International Relations (GSIR), International University of Japan (IUJ), e-mail: yjinnai@iuj.ac.jp.

1 Introduction

School accountability has been a centerpiece of public education reform in the United States for the last two decades. Accountability programs typically evaluate schools based on student achievement on statewide standardized tests. With a variety of rewards, including cash bonuses, for high-performing schools and teachers as well as sanctions for low-performing ones, accountability programs have increased student achievement as policy makers anticipated (Ladd, 1999; Hanushek and Raymond, 2005; Jacob, 2005; Figlio and Rouse, 2006).[1]

From a theoretical point of view, teacher performance pay programs are designed to solve at least two problems: screening and moral hazard. Screening involves how to select effective teachers. Although teacher quality is one of the most important determinants of student achievement, it is hard to identify and hire only productive teachers.[2] Moral hazard arises from the fact that teachers' actions in classrooms are hidden from their supervisors, and so education authorities cannot force them to exert sufficient effort. In order to overcome these screening and monitoring problems, performance pay programs provide teachers with financial incentives based on student achievement, expecting to remove low-performing teachers and to induce more effort from teachers.[3]

Empirically, a series of non-experimental studies in Israel (Lavy, 2002, 2009) and experimental studies in Kenya and India (Glewwe, Ilias and Kremer, 2010; Muralidharan and Sundararaman, 2011) have found that the introduction of monetary incentives for teachers significantly increased teacher effort and student achievement as expected. However, the impact of teacher performance pay programs in the U.S. is less clear. Although Figlio and Kenny (2007) document a positive relationship between individual-based teacher incentives

[1]Cullen and Reback (2006), Figlio and Getzler (2006), and Figlio and Loeb (2011) provide detailed reviews of the relationship between accountability programs and student achievement as well as of the gaming behaviors of schools under accountability pressure.

[2]Rockoff (2004), Rivkin, Hanushek and Kain (2005), Kane, Rockoff and Staiger (2008), Kane and Staiger (2008), and Rockoff et al. (2008) discuss the relationship between observable teacher characteristics and their added value.

[3]See Holmstrom and Milgrom (1991) and Neal (2011) for the theoretical framework of the multi-tasking principal-agent model and the optimal design of incentive schemes for educators.

and student achievement, there is no evidence that teacher incentives increased student achievement in New York City (Goodman and Turner, 2010; Fryer, 2011). By contrast, incentives in North Carolina affected teachers differently depending on their prior performance (Vigdor, 2008; Ahn and Vigdor, 2014; Jinnai, 2014). In addition to these mixed empirical findings, little is known about the potentially negative effects of removing such incentives.

In this study, I address the effects of repealing a teacher incentive program by exploiting the novel experience of North Carolina, where the state government abolished its bonus system in 2009. Beginning from the 2005-06 school year, each qualified teacher was awarded a school-wide bonus up to $1,500 per person each year. In 2007-08, however, the state government reduced the maximum amount to $1,053. As a result of continuing economic downturn, the state repealed the incentive program entirely in 2008-09. These policy changes were exogenous because the state government announced its new guidelines at the end of each school year, and therefore no teachers were able to expect such amendments in advance.

An empirical challenge of estimating the impact of repealing the incentive program stems from the fact that all public schools across the state were affected by the new policy, leaving no control group to aid in the identification of treatment effects. Accordingly, this paper follows Neal and Schanzenbach (2010) and compares student achievement in post-reform years with predicted achievement levels that students would have reached had the state maintained its initial bonus program.[4] The difference between realized and predicted values helps assess the impact of *reducing* and *repealing* incentive bonuses, respectively.

Estimation results show that, after the *reduction* in incentive bonuses, student achievement significantly decreased at the lowest-performing schools, which had consecutively failed to receive bonuses. Moreover, after the *repeal* of the performance pay program, student achievement further decreased not only at the lowest-achieving schools but also at the second-lowest-achieving schools. On the other hand, the highest-performing schools, which

[4]Neal and Schanzenbach (2010) estimate the effects of introducing an accountability program on student achievement in Chicago, and conclude that the accountability induced schools to focus on students near the middle of the achievement distribution. Using a similar approach, Ladd and Lauen (2010) find little evidence that failing schools in North Carolina ignore those students far below proficiency.

had successively qualified for bonuses, maintained their high standard even after the policy changes.

These results are counterintuitive because it is high-performing schools that should expect negative impact as a result of the new policies; low-performing schools would not have expected to qualify for bonuses even if the state had kept the incentive program. I provide two possible explanations for these striking results by examining the differences in teacher characteristics between low- and high-achieving schools. One is that inexperienced teachers are more likely to be hired by low-performing schools and earn less than teachers at high-performing schools who are more experienced; thus, the amount of the bonuses makes up a larger proportion of their annual salary and provides a greater incentive. The other explanation is that teachers in high-achieving schools are more self-disciplined and do not only work for monetary rewards.

This paper makes a three-fold contribution to the literature on school accountability and teacher incentive programs. First and most importantly, this is the first study that exploits exogenous policy changes that reduced and then repealed teacher incentives. The findings are informative particularly today when states and countries frequently cut their education budgets. Second, this study investigates the differential effects of repealing cash bonuses by measuring changes in student achievement at different points in prior performance, suggesting the need for a better design of incentives for schools at each point. Third, in contrast to other related studies, this paper uses the data from the sophisticated accountability program of North Carolina, which was established well before the No Child Left Behind Act of 2001 (NCLB).

The rest of the paper is organized as follows. The next section describes North Carolina's accountability program. Section 3 details data on students, teachers, and schools. Section 4 illustrates the empirical framework, Section 5 demonstrates estimation results, and Section 6 concludes.

2 North Carolina's bonus program

North Carolina has had a carefully designed educational accountability system in place since the 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.[5] 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 that judge schools on the average level of test scores, although the level model is the basis for the federal NCLB and many other accountability programs.

Beginning from 2005-06 school year, each school's average growth across all students in all subjects is calculated.[6] On the basis of average growth scores, the state government classifies its public schools into one of the three categories: High Growth, Expected Growth, or Less than Expected. For elementary and middle schools that have met High Growth, all of their certified teachers receive the same amount of $1,500 in bonuses per person each year. For the schools that have met Expected Growth, certified teachers receive $750. By contrast, teachers at Less than Expected schools receive none.

In 2007-08, however, the amount of bonus was reduced to $1,053 at High Growth schools and to $527 at Expected Growth schools. 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. Since 2008-09, incentive bonuses have been suspended because of the state's economic condition.

[5] The background of North Carolina's accountability program is described in Appendix A.1.

[6] The evaluation formula is shown in Appendix A.2.

3 Data

In this study, I combine data from two sources. Detailed data sets on students, teachers, and schools are provided by the North Carolina Education Research Data Center (NCERDC), and each school's average growth score, the main outcome variable, is provided by the North Carolina Department of Public Instruction (DPI).

I use data for school years from 2005-06 (year 2006 hereafter) to 2009-10 (year 2010) because the new growth formulas implemented in 2006 render comparisons to previous years inappropriate. Since non-regular schools and high schools follow different rules for accountability, I further restrict the sample to public regular elementary and middle schools, resulting in a sample of approximately 1,800 schools each year.

Table 1 shows the descriptive statistics (means and standard deviations) for public schools in North Carolina from 2006 to 2010. Almost all the school characteristics are constant over time; enrollment is around 550 with class size of 20 and student-teacher ratio of around 15. Of the students, 53-54% are white, 29-33% are black, and 51-57% are eligible for free or reduced-price lunch programs. The proportion of Hispanic students increased from 9.1% in 2006 to 11.9% in 2010. Regarding teachers, teaching quality has improved over time; 96.6% of teachers had licenses in 2010 (up from 89.7% in 2006), whereas 27.2% had advanced degrees (Master's or higher, up from 24.9%). The proportion of teachers with experience of less than three years decreased from 24.1% to 19.6%, while the turnover rate fell from 21.4% to 11.5%.

As described in Section 2, each public school in North Carolina is classified into one of three categories associated with different amounts of bonuses by the two variables of academic achievement: average growth and change ratio. Since each student's growth score is calculated as in equation (6) with a discount factor δ being less than one, the mean of the average growth scores across all schools can be greater than zero each year. In fact, as shown in Table 1, average growth increased from 0.002 in 2006 to 0.180 in 2008 and then dropped; change ratio shows a similar pattern.

Table 1: Descriptive statistics of public schools in North Carolina from 2006 to 2010

	2006	2007	2008	2009	2010
School characteristics					
Enrollment	545.1	555.4	557.8	544.8	546.4
	(240.3)	(242.3)	(242.9)	(242.8)	(233.0)
Class size	19.9	19.7	20.0	20.1	20.6
	(2.7)	(2.7)	(2.5)	(3.7)	(3.4)
Student-teacher ratio	14.9	14.9	14.8	14.5	15.0
	(4.4)	(4.7)	(2.5)	(2.7)	(2.3)
Student characteristics					
% White	0.544	0.539	0.531	0.527	0.527
	(0.286)	(0.286)	(0.287)	(0.285)	(0.283)
% Black	0.329	0.289	0.286	0.320	0.315
	(0.258)	(0.252)	(0.252)	(0.250)	(0.243)
% Hispanic	0.091	0.101	0.109	0.114	0.119
	(0.094)	(0.103)	(0.109)	(0.112)	(0.115)
% Free lunch eligible	0.521	0.510	0.518	0.530	0.565
	(0.212)	(0.208)	(0.216)	(0.221)	(0.219)
Teacher characteristics					
% License	0.897	0.947	0.947	0.957	0.966
	(0.099)	(0.076)	(0.071)	(0.064)	(0.054)
% Advanced degree	0.249	0.254	0.256	0.264	0.272
	(0.095)	(0.096)	(0.097)	(0.098)	(0.103)
% Experience with less than 3 years	0.241	0.234	0.236	0.221	0.196
	(0.106)	(0.109)	(0.106)	(0.107)	(0.103)
% Experience with 3-10 years	0.280	0.284	0.291	0.299	0.315
	(0.085)	(0.088)	(0.088)	(0.090)	(0.095)
% Experience more than 10 years	0.482	0.482	0.476	0.480	0.489
	(0.128)	(0.131)	(0.129)	(0.128)	(0.130)
% Turnover	0.214	0.220	0.127	0.122	0.115
	(0.103)	(0.103)	(0.069)	(0.069)	(0.066)
Academic achievement					
Average growth	0.002	0.050	0.180	0.105	0.124
	(0.103)	(0.107)	(0.146)	(0.108)	(0.103)
Change ratio	1.079	1.265	2.046	1.551	1.630
	(0.364)	(0.429)	(1.467)	(2.447)	(0.607)
N	1814	1822	1807	1831	1828

Note: Each column reports the mean and the standard deviation (in parentheses) of the school-level variables from different academic years. The number of schools is reported at the bottom of the table.

4 Empirical framework

A difficulty of estimating the impact of repealing a bonus incentive program is that the incentives were removed simultaneously across the state, leaving no reliable control group to aid in the identification of treatment effects. In order to credibly compare school performance before and after the repeal of the incentive system, this study follows Neal and Schanzenbach (2010), who consider the difference between realized student achievement and predicted achievement to estimate the effects of introducing an accountability program in Chicago.

I divide the sample into three categories: pre-reform (2006 to 2008), mid-reform (2009), and post-reform (2010). During the pre-reform years, teachers taught classes, expecting a full bonus of $1,500. Their expectations fell to $1,053 in the mid-reform and to zero in the post-reform year, as described in Section 2. The basic idea in identifying the impact of *reducing* and *repealing* bonus incentives is to predict mid-reform outcomes (which schools could have achieved had the state maintained the full bonus program) by using the pre-reform data, and to predict post-reform outcomes by using the mid-reform data. The estimation procedure is as follows:

Step 1. With only the pre-reform data, run a regression of the form

$$y_{st}^{pre} = X_{st}^{pre}\beta + \theta_s + \epsilon_{st}, \tag{1}$$

where y_{st}^{pre} is the average growth score of school s in year t $(= 2006, 2007, 2008)$. X_{st}^{pre} denotes school inputs that include school characteristics (e.g., total enrollment and class size), student proportions (e.g., race and free lunch eligibility), and teacher quality (e.g., teacher experience and advanced degree holders).[7] ϵ_{st} is an i.i.d. error term. I run the regression with and without school fixed effects θ_s.

Step 2. Calculate each school's predicted growth score in the mid-reform year ($t = 2009$)

[7]For each period t and each school s, X_{st} includes all variables of school, student, and teacher characteristics listed in Table 1.

as follows, using the estimates of β and θ_s in Step 1:

$$\widehat{y}_{st}^{mid} = X_{st}^{mid}\widehat{\beta} + \widehat{\theta}. \tag{2}$$

Note that in equation (2), in order to predict the mid-reform outcomes under the counterfactual full bonus program, the mid-reform inputs X_{st}^{mid} are used, while $\widehat{\beta}$ and $\widehat{\theta}$ are drawn from the previous step.

Step 3. Divide the sample into four groups, as in Table 2, according to each school's bonus receipt from the past two years. Group 1 consists of schools that did not receive bonuses in both 2007 and 2008. Group 2 schools are those that received bonuses in 2007 but not in 2008, while group 3 schools received bonuses only in 2008. Schools in group 4 qualified for bonuses in both years.

Table 2: The classification of schools based on prior bonus receipt

	Bonus receipt	
	2007	2008
Group 1	No	No
Group 2	Yes	No
Group 3	No	Yes
Group 4	Yes	Yes

Step 4. For each of the four groups, take the difference between the realized mid-reform outcomes and the predicted values as

$$d_{st}^{mid} = y_{st}^{mid} - \widehat{y}_{st}^{mid}, \tag{3}$$

and then calculate the average of d_{st}^{mid} to estimate the differential effects of reducing bonus incentives on school performance.

In order to evaluate the impact of completely removing the incentives in 2010, I follow modified estimation steps. Since the mid-reform period consists of a single year of 2009, Step 1 cannot include school fixed effects θ_s in equation (1). Thus, I employ the following two different methods: (i) an OLS estimation of the form

$$y_{st}^{mid} = X_{st}^{mid}\beta + \epsilon_{st}, \tag{4}$$

using only the data from mid-reform year ($t = 2009$); and (ii) a school-fixed-effects estimation of the form

$$y_{st}^{pre\_mid} = X_{st}^{pre\_mid}\beta + \theta_s + \epsilon_{st}, \tag{5}$$

using the data from $t = 2008, 2009$.

Although the second method uses the data also from 2008, not the mid-reform but the pre-reform year, it takes advantage of employing school fixed effects. In addition, teachers received bonuses up to $1,053 even in the mid-reform period, which is far from none in the post-reform period. Therefore, this paper does not distinguish between the differences in bonus size in $t = 2008$ and 2009, and I prefer the estimation results from the second method.

5 Results

5.1 Impact of reducing incentive bonuses

Table 3 presents, for each of the four groups, the mean differences between actual average growth scores in 2009 (when teachers expected reduced bonuses) and predicted scores that schools could have gained had the state kept the full bonus system. The first column shows the average differences calculated by using the estimate of β from the pooled OLS estimation results in Step 1, with the data from 2006 to 2008. The second column shows those calculated by using the estimates of β and θ_s from the school-fixed-effects estimation results in Step 1 with the same sample.

The baseline results of the first column demonstrate that in 2009 the mean of the average growth scores of group 1 schools (which did not receive bonuses in both 2007 and 2008) was 0.076 standard deviations (s.d.) lower than predicted, with significance at the 1% level. Although the magnitude was smaller, group 2 schools (which had qualified for bonuses only in 2007) also had average growth scores 0.038 s.d. lower than predicted; group 3 schools (which only qualified for bonuses in 2008) yielded 0.027 s.d. lower than predicted. By contrast, group 4 schools (which had received bonuses both in 2007 and 2008) achieved 0.019 s.d. higher than predicted.

Table 3: The effects of reducing incentive bonuses on school performance in 2009

	(1) OLS with data from 2006 to 2008	(2) Fixed Effects with data from 2006 to 2008
Group 1	-0.076***	-0.049**
($N = 66$)	(0.013)	(0.023)
Group 2	-0.038*	-0.053
($N = 35$)	(0.019)	(0.042)
Group 3	-0.027***	0.000
($N = 332$)	(0.005)	(0.009)
Group 4	0.019***	0.048***
($N = 883$)	(0.003)	(0.005)

Note: Each column reports the mean difference between realized and predicted average growth scores, with standard errors in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Group 1 schools qualified bonuses in neither 2007 nor 2008; group 2 received bonuses only in 2007; group 3 only in 2008; group 4 in both years.

Even after controlling for school fixed effects, the results in the second column illustrate a similar pattern. The growth of group 1 school was, on average, significantly lower (0.049 s.d.) than predicted, while that of group 4 schools was significantly higher (0.048 s.d.). On the other hand, realized growth scores were not significantly different from predicted values for group 2 and 3 schools. These results suggest that reducing the amount of bonuses from $1,500 to $1,053 per teacher per year affected the lowest-performing schools more negatively than the highest-performing ones.

5.2 Impact of repealing incentive bonuses

Table 4 presents, for each of the four groups, the difference between actual average growth scores in 2010 (when teachers expected no bonuses) and predicted scores that schools could have achieved. Similar to Table 3, the first column shows the results from the OLS estimation in Step 1, and the second column shows those from the school-fixed-effects estimation. The baseline OLS results again demonstrate that the lowest-performing schools (group 1) scored lower than predicted and that the highest-performing schools (group 4) scored higher than predicted.

Table 4: The effects of repealing incentive bonuses on school performance in 2010

	(1) OLS with data from 2009	(2) Fixed Effects with data from 2008 and 2009
Group 1	-0.003	-0.115***
($N = 61$)	(0.014)	(0.029)
Group 2	-0.013	-0.084***
($N = 166$)	(0.009)	(0.015)
Group 3	0.045***	-0.018
($N = 69$)	(0.013)	(0.016)
Group 4	0.022***	0.005
($N = 1327$)	(0.003)	(0.004)

Note: Each column reports the mean difference between realized and predicted average growth scores, with standard errors in parentheses. *** denotes significance at the 1% level. Group 1 schools qualified bonuses in neither 2008 nor 2009; group 2 received bonuses only in 2008; group 3 only in 2009; group 4 in both years.

The preferred results from the fixed-effects estimation in the second column illustrate a similar pattern to those in Table 3. Schools in group 1 and 2 scored lower than predicted, but in this case their magnitudes are larger, leading to worse achievement. Group 1 schools yielded, on average, growth scores that were 0.115 s.d. lower than predicted, while group 2 schools achieved 0.084 s.d. lower; both are significant at the 1% level. These results again suggest that repealing the bonus incentives was more detrimental to lower achieving schools than to higher achieving ones.

5.3 Discussion

The findings shown in previous subsections are somewhat surprising. After reducing or repealing incentive bonuses, it is high-performing schools that should expect a negative shock, because low-performing schools are unlikely to expect to become qualified for bonuses even under incentive programs. However, the results show that it is low-performing schools that suffer from lower-than-expected growth scores. In this subsection, I provide two potential reasons why the policy changes aggravated low-achieving schools.

Table 5: Teacher characteristics of each group in 2010

	(1) Group 1	(2) Group 2	(3) Group 3	(4) Group 4	(5) G1 - G4
% License	0.924	0.956	0.930	0.971	-0.048***
	(0.076)	(0.068)	(0.086)	(0.046)	(0.005)
% Advanced degree	0.235	0.243	0.236	0.279	-0.044***
	(0.103)	(0.094)	(0.096)	(0.103)	(0.013)
% Experience with less than 3 years	0.266	0.233	0.222	0.185	0.081***
	(0.103)	(0.103)	(0.111)	(0.099)	(0.013)
% Experience with 3-10 years	0.329	0.309	0.345	0.313	0.016
	(0.094)	(0.094)	(0.095)	(0.095)	(0.012)
% Experience more than 10 years	0.406	0.458	0.433	0.503	-0.097***
	(0.113)	(0.114)	(0.121)	(0.128)	(0.017)
% Turnover	0.166	0.134	0.139	0.109	0.057***
	(0.105)	(0.080)	(0.066)	(0.059)	(0.008)
N	67	179	83	1436	–

Note: The first through the fourth columns report the mean and the standard deviation (in parentheses) of teacher characteristics for each group. Group 1 schools did not qualify for bonuses in both 2008 and 2009; group 2 received bonuses only in 2008; group 3 only in 2009; group 4 in both years. The number of schools is reported at the bottom of the table. The fifth column reports the differences in means between group 1 and group 4, with standard errors in parentheses. *** denotes significance at the 1% levels.

The first explanation involves the size of bonuses. Although the full bonus amount of $1,500 may not be large enough for teachers to exert additional effort, it consists of approximately 3.5% of their average annual salary.[8] Since teacher salary depends on experience in profession, the fixed amount of incentive bonuses are more attractive to teachers with

[8] DPI reports that the statewide average salary of teachers for fiscal year of 2010-11 is $42,416 without benefits.

less experience. Therefore, the bonus program has a potentially larger impact on those inexperienced teachers.

Table 5 compares the means of teacher characteristics for each of the four groups. The first column shows that, on average, 26.6% of teachers at group 1 (the lowest-performing) schools have experience less than three years, while the fourth column shows that the figure is 18.5% in group 4 (the highest-performing) schools; the difference of 8.1 percentage points, shown in the fifth column, is significant at the 1% level. On the other hand, 40.6% of teachers at group 1 schools have experience more than ten years, while this proportion is 50.3% for group 4 schools; the difference is again significant at the 1% level. As a result of a relatively higher ratio of inexperienced teachers, group 1 schools could have been negatively affected by the reduction and the repeal of incentive bonuses.

The second possible explanation is the quality of teachers. Table 5 also illustrates that the proportions of teachers with licenses and advanced degrees are lower in group 1 than in group 4. Moreover, turnover rate is higher in group 1 schools. All of these differences, with significance at the 1% level, indicates that teachers at high-achieving schools are more qualified and possibly better self-disciplined to work not for cash bonuses but for the sake of their students.

As Fryer (2011) describes, teacher performance pay programs can lead to unexpected results.[9] However, the above two interpretations will be able to explain why lower achieving schools were more negatively affected by the policy changes in the bonus program compared to their higher performing counterparts.

[9]Fryer (2011) finds no evidence that teacher incentives increased student achievement in New York, and provides four explanations: (i) incentives may not be large enough; (ii) the incentive scheme is too complex; (iii) group-based incentives may not be effective; and (iv) teachers may not know how they can improve student achievement.

6 Conclusion

As a growing number of states and countries have introduced school accountability programs and teacher performance pay systems, recent studies have extensively focused on assessing the effectiveness of such programs. However, no existing research has yet examined the impact of repealing teacher incentive bonuses.

This paper exploits novel data from North Carolina, where the state government reduced and then repealed its teacher incentive bonuses due to its poor economic condition. By comparing the realized achievement and predicted values, the study finds that both reducing and repealing incentives negatively affects low-performing schools. After the reduction in bonuses, the average growth of the lowest-performing schools significantly decreased by 0.049 s.d. After the repeal of the incentive program, the figure further decreased by 0.115 s.d. at those lowest-achieving schools as well as by 0.084 s.d. at the second-lowest-achieving schools. On the other hand, high-achieving schools maintained their high standard even after the policy changes.

These results are counterintuitive because it is high-performing schools that should expect negative impact from the new policies. Based on the differences in teacher characteristics between high- and low-achieving schools, I provide two potential reasons for the negative impact on lower performing schools. One is that those schools have a larger proportion of inexperienced teachers, to whom the fixed size of bonuses is more appealing. As a result, the repeal of the bonuses represents a relatively larger drop in their salaries. The other explanation is that teachers at higher performing schools are qualified and self-disciplined enough not to work only for incentive bonuses, and thus are not affected by the policy changes.

The empirical findings from this study are informative to the climate today, wherein states and countries frequently cut their education budgets. One particularly important policy implication is that once teacher incentives are implemented, reducing or repealing them is not cost-free. Therefore, states and countries must carefully consider whether or not to introduce these programs in the first place. If the repeal of the programs cannot be avoided, it seems reasonable to remove incentives from high-performing schools first, since this study clearly shows that the lowest-performing schools are those that are most negatively affected by such policy changes.

References

Ahn, T., and J. Vigdor. 2014. "When incentives matter too much: Explaining significant responses to irrelevant information." *Working Paper at the University of Kentucky*, 1–44.

Cullen, J., and R. Reback. 2006. "Tinkering toward accolades: School gaming under a performance accountability system." In *Advances in Applied Microeconomics*. . 14 ed., , ed. T. Gronberg and D. Jansen, Chapter 1, 1–34. Emerald Group Publishing Limited.

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

Figlio, D., and L. Getzler. 2006. "Accountability, ability, and disability: Gaming the system?" In *Advances in Applied Microeconomics*. , ed. T. Gronberg and D. Jansen, Chapter 2, 35–49. Emerald Group Publishing Limited.

Figlio, D., and L. Kenny. 2007. "Individual teacher incentives and student performance." *Journal of Public Economics*, 91(5-6): 901–914.

Figlio, D., and S. Loeb. 2011. "School accountability." In *Handbook of the Economics of Education*. Vol. 3, , ed. E. Hanushek, S. Machin and L. Woessmann, Chapter 8, 383–421. North Holland.

Fryer, R. 2011. "Teacher incentives and student achievement: Evidence from New York City public schools." *NBER Working Paper No.16850.*

Glewwe, P., N. Ilias, and M. Kremer. 2010. "Teacher incentives." *American Economic Journal: Applied Economics*, 2(July): 205–227.

Goodman, S., and L. Turner. 2010. "Teacher incentive pay and educational outcomes: Evidence from the New York City bonus program." *Working Paper at Columbia University.*

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

Holmstrom, B., and P. Milgrom. 1991. "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics, and Organization*, 7: 24–52.

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.

Jinnai, Y. 2014. "The effects of a teacher performance-pay program on student achievement: A regression discontinuity approach." *Working Paper at International University of Japan.*

Kane, T., and D. Staiger. 2008. "Estimating teacher impacts on student achievement: An experimental evaluation." *NBER Working Paper No.14607.*

Kane, T., J. Rockoff, and D. Staiger. 2008. "What does certification tell us about teacher effectiveness? Evidence from New York City." *Economics of Education Review*, 27(6): 615–631.

Ladd, H. 1999. "The Dallas school accountability and incentive program: an evaluation of its impacts on student outcomes." *Economics of Education Review*, 18: 1–16.

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.

Lavy, V. 2002. "Evaluating the effect of teachers' group performance incentives on pupil achievement." *Journal of Political Economy*, 110(6): 1286–1317.

Lavy, V. 2009. "Performance pay and teachers' effort, productivity, and grading ethics." *American Economic Review*, 99(5): 1979–2011.

Muralidharan, K., and V. Sundararaman. 2011. "Teacher performance pay: Experimental evidence from India." *Journal of Political Economy*, 119(1): 39–77.

Neal, D. 2011. "The design of performance pay in education." In *Handbook of the Economics of Education*. Vol. 4, , ed. E. Hanushek, S. Machin and L. Woessmann, Chapter 6, 495–550. North Holland.

Neal, D., and D. Schanzenbach. 2010. "Left behind by design: proficiency counts and test-based accountability." *Review of Economics and Statistics*, 92(May): 263–283.

Rivkin, S., E. Hanushek, and J. Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica*, 73(2): 417–458.

Rockoff, J. 2004. "The impact of individual teachers on student achievement: Evidence from panel data." *American Economic Review*, 94(2): 247–252.

Rockoff, J., B. Jacob, T. Kane, and D. Staiger. 2008. "Can you recognize an effective teacher when you recruit one?" *NBER Working Paper No.14485*.

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

Appendices

A Accountability program in North Carolina

A.1 Background

In 1996, the North Carolina State Board of Education (SBE) developed a school accountability program, referred to as the ABCs of Public Education, which focused on strong **A**ccountability, teaching the **B**asics with an emphasis on high educational standards, and maximum local **C**ontrol. In 2002-03, the ABCs program was expanded to incorporate the new statutory accountability requirements of the federal NCLB. In 2005-06, new growth formulas were implemented that make comparisons to previous years inappropriate.

The ABCs accountability program sets growth and performance standards for each elementary, middle, and high school in the state. End-of-Grade (EOG) and End-of-Course (EOC) test results, and other selected components, are used to measure a school's growth and performance. Schools that attain the standards are normally eligible for incentive awards or other recognition. Schools where growth and performance fall below specified levels are designated as low-performing, and may receive mandated assistance based on action by the SBE.

A.2 Evaluation

In 2005-06, new formulas for determining student performance were introduced primarily to account for reversion to the mean. The new formula calculates each student's academic growth, using standardized test scores for each grade and each year. In practice, student i's growth in year t at grade g is calculated as:

$$growth_{igt} = Z_{igt} - \delta\Big(\frac{Z_{i,g-1,t-1} + Z_{i,g-2,t-2}}{2}\Big), \tag{6}$$

where Z_{igt} is a normalized test score based on the mean and standard deviation from the first year a particular test was used in the state. The discount factor δ accounts for mean reversion: $\delta = 0.92$ when two-year observation is available, and 0.82 when only a single year is available.

For each school, its average growth across all students in all subjects is calculated. In elementary and middle schools, if a school's average growth is equal to or greater than zero, the school is said to have met "Expected Growth", and all of its certified teachers receive the same amount of $750 bonuses per person per year. If a school met expected growth (i.e., "average growth" ≥ 0) and at least 60% of its students achieve the required growth (which is defined as "change ratio" $\geq \frac{60\%}{(100-60)\%} = 1.5$), the school is said to have met "High Growth"; teachers are eligible for $1,500 bonuses. Thus, schools that achieve strong test score growth by raising the performance of a limited number of students generally do not receive the full bonus.

A.3 Definition of Expected Growth

A school's ABCs growth status is determined by its growth calculation and its change ratio (a measure of the percent of students meeting their individual growth targets). A school's grade span and/or courses determine the composition of these measures, as described below. The average growth for a school may include:

(1) Average growth on EOG reading and mathematics for grades 3-8 and any EOC tests.

(2) Change over a two-year baseline in the percent of students completing the college/university prep and college tech prep courses of study.

(3) Change in the competency passing rate.

(4) Change in the ABCs dropout rate.

The schools whose average growth is equal to the growth expectation (shown by an average difference of 0.00 or better) are said to have met Expected Growth.

A.4 Definition of High Growth

The change ratio used to determine the attainment of high growth is calculated as follows. The factors are arranged such that the number of students meeting their individual growth standards is in the numerator along with the change in competency pass rate and college/university prep and college tech prep courses of study. Students not meeting their individual growth standard are in the denominator and the decrease in dropout rate is subtracted from the denominator. Schools that have an average growth of 0.00 or better (met expected growth) and have a change ratio of 1.50 or better are said to have met High Growth.