The Missing Link: Estimating the Impact of Incentives on Effort and Effort on Production Using Teacher Accountability Legislation

Tom Ahn†
Duke University
October 13, 2008

Abstract

Teacher effort, a critical component of education production, has been ignored in the literature due to measurement difficulties. I use a principal-agent model, data on NC public elementary schools, and the state’s unique accountability system that rewards teachers for year-over-year school-level academic growth, to distill effort from absence data and capture its effect on student achievement. I find low effort at low and high probabilities of bonus receipt, high effort when the bonus outcome is in doubt, and free-ridership. Teachers respond optimally to incentives, effort strongly impacts achievement, and the effect varies across racial groups. Policy changes that 1) restrict the bonus criterion to classroom-level to solve the free-rider problem, 2) track minority scores to close the racial achievement gap, and 3) strengthen the bonus criterion to induce effort, all back-fire due to the non-linear effect of marginal teacher effort on achievement, resulting in lower statewide scores.

Keywords:
JEL I21

*I thank Peter Arcidiacono, Pat Bayer, Helen Ladd, Tom Nechyba, Stephen Ryan, Justin Trogdon, and Jacob Vigdor. I am grateful to the Center for Child and Family Policy for access to the NCDPI dataset and to the Spencer Foundation for financial support. All remaining errors are my own.
†Center for Child and Family Policy, Duke University, Durham NC 27708; thomas.ahn@duke.edu
1 Introduction

With decades of research attempting to formulate and estimate an education production function, the surprising consensus so far has been that observable school and teacher characteristics fail to account for large variations in achievement growth (Rivkin, Hanushek, and Kain (2005)). In response, economists have cast a wider net, looking at peer effects, school choice and financing, discipline policy, and ability tracking among other factors to identify the determinants of student achievement (Arcidiacono et.al. (2007), Hoxby and Weingarth (2005), Lazear (2001), Nechyba (2003), among many others).

In spite of this far-ranging search among the more exotic components of education production, there is agreement among those involved in the education process that teacher quality is one of the most important determinants of a student’s success in his academic career. The literature has identified teaching experience, schooling, and credentials as having a positive effect on student outcome (See Goldhaber and Anthony (2007), Rockoff (2004), Clotfelter, Ladd and Vigdor (2007), among others). The identified qualities of teachers are, for the most part, exogenous ability characteristics, and they fail to account for a majority of the variation in student achievement. Most of the unaccounted-for effect of teachers has been called unobservable ability and modeled as individual-specific constants in fixed-effects models (Rockoff (2004), Rivkin, Haushek, and Kain (2005)).

This paper explores an important and unexplored dimension of teacher quality that drives student achievement growth: effort. The current political debate of increasing the role of merit-pay, scrapping of traditional fixed pay scale, and doing away with job security are aimed at inducing effort, the endogenous and manipulable characteristic common to teachers. Policy makers have used accountability legislation as a blunt instrument to more efficiently motivate teachers by introducing market forces. Federal legislation such as the No Child Left Behind Act (NCLB) and state legislations have attempted to raise student achievement and close the racial achievement gap.

This issue can be understood as a variant of the principal-agent problem (See Sappington (1991) for a review). Usually, the innate ability of the agent is outside the principal’s control, and his/her role is to design a contract to maximize the effort of the agent(s). For this study, the principal (government) is setting rewards for a cooperative team project (school-wide
achievement) in which the quality of agents (teachers) is exogenous and heterogeneous, the output signal (student achievement) is noisy, and an imperfect public signal of the agents’ effort is available (teacher absence taking behavior). I propose a method of measuring effort using a teacher incentive bonus system, analyze the intended and unintended effect of using accountability legislation to influence effort (efficiency as well as distributional implications), and its eventual result on overall student achievement and the racial achievement gap.

While there is considerable debate on whether accountability systems help students as intended, it is clear that teachers and administrators respond to incentives. In fact, a sizable literature has examined the unintended, sometimes perverse responses of teachers to poorly formed accountability systems. Other studies have shown that mishandled incentive programs do not result in the desired outcome. Teachers shift resources away from low performing students to students with a higher probability of reaching the required threshold (Neal and Schanzenbach (2007), Booher-Jennings (2005), and White and Rosenbaum (2007)). These examples are not necessarily indictments of accountability programs, but warnings that setting arbitrary standards and expecting good results is naïve. Studies have shown that performance incentives for teachers can lead to higher academic performance (Figlio and Kenny (2006), Clotfelter and Ladd (1996), Ladd (1999), Eberts, Hollenbeck, and Stone (2002)). The accumulated evidence points to the fact that teachers have a great deal of influence on student outcomes, and incentive-laden policy can be effective in influencing teacher behavior.

However, there are several open questions that need to be addressed. Previous studies have used student achievement as signals for changes in teacher behavior, but there has been no systematic attempt to look at the causal link from accountability incentives to teacher effort to student achievement. There is very little evidence on how many dollars are required to raise teacher effort by enough to affect student performance by a targeted amount. This study will provide a structural model to estimate the causal effects. Furthermore, the focus of research to date has been on the distributional changes in achievement. Besides the change in distribution of resources/effort, which clearly defines winners and losers among the student

---

1 Examples of these behaviors include classifying marginal students as disabled or suspending them to prevent testing (Cullen and Reback (2002), Figlio and Getzler (2002), Jacob (2005), Figlio (2006)), ‘teaching to the test’ (Carnoy and Loeb (2002), Grissmer and Flanagan (1998), Hanushek and Raymond (2004)), and cheating by teachers who alter students’ answer sheets (Jacob and Levitt (2003)).

2 Note that these studies imply a reallocation of resources, not an increase or decrease in effort.
population, there can still be net positive or negative effects, depending on aggregate effort level changes due to incentive policy. This study will examine efficiency as well as distributional implications of accountability policies and their effect on incentives and performance.

I use a unique dataset collected in the state of North Carolina that tracks the academic history, demographic information, and teacher and peer exposure of all students in the public school system. This dataset, along with the unique teacher incentive system that pays out cash bonuses for school-level year-over-year improvement in student test scores, makes it possible to observe the teacher’s level of effort, identify her effort response to incentives, gauge the effect of effort on student achievement in general and in closing the racial achievement gap in particular, and evaluate the efficacy of the accountability policy.

The theoretical results show that the key incentive variables of interest are the number of teachers at a school and the probability of bonus receipt. The number of teachers at the school determines the magnitude of the free-rider (incentive dilution) effect present. As the number of teachers increases (while the size of individual rewards stays constant), each teacher’s optimal effort declines, ceteris paribus. The probability of the bonus receipt determines the marginal effect of additional effort by a teacher. The non-linearity in the probability of bonus receipt translates to low effort exertion when there is very little or very high probability of bonus receipt (because a single teacher’s effort fails to substantially change the bonus receipt likelihood) and high effort exertion when the bonus receipt is in doubt (when a teacher’s effort matters).

The econometric results show that teacher effort is an important component of student achievement, white students are more sensitive to teacher effort input compared to minority students, and teacher adjust their efforts to maximize expected benefits. Using parameter estimates from the model, I simulate policy changes that 1) change the bonus regime from cooperative (school-wide average criterion) to non-cooperative (teacher individual class average criterion), 2) make bonus receipt dependent upon minority achievement, and 3) make the criterion tougher to reach. The first policy change tries to reduce the free-rider problem intrinsic to the incentive policy; the second change attempts pay more attention to traditionally disadvantaged groups; the third change aims to induce higher teacher effort by making the criterion more difficult to reach.

The simulation results show that the policy changes, all first-order attempts to increase
achievement, end up lowering average achievement. While the first policy change does reduce
the general free-rider effect, it also makes the bonus receipt probabilities much more stark,
firmly dividing teachers between those who will get the bonus for certain, and those who will
not get the bonus. Both groups of teachers are induced to exert low effort. The second policy
change reduces average effort across schools due to the lower sensitivity to teacher effort
of minority students. The third policy change makes almost all teachers pull back effort,
indicating that the status quo state standard was already too difficult to get teachers to give
a high level of effort. While the second policy narrows the racial achievement gap somewhat,
the other two policies increase the gap. This big lesson is that setting arbitrary performance
standards and hoping for the best is naïve.

The next section details the North Carolina accountability system. Section 3 presents a
simple theoretical model. Section 4 describes the data, and section 5 introduces the econo-
metric model. In section 6, I present the results. Section 7 uses the parameters estimates and
discusses the policy implications. I conclude in section 8. All proofs, tables, and figures are
in the appendix.

2 The North Carolina Accountability System

The North Carolina accountability program (also known as ABC) began in the 1995-96 aca-
demic year. While the system has grown in complexity and has gone through minor alterations
in the details of execution,³ the principal mechanism of offering cash incentives for student
achievement gains has remained unchanged for more than a decade.

North Carolina public school students in grades 3 through 8 must take End-of-Grade exams
in reading and mathematics. The test is on a developmental scale, allowing comparison of
scores from consecutive grades. Using the formula defined below, North Carolina Department
of Public Instruction (NCDPI) determines the required achievement gains for each school based
on the school’s students’ performance last year on the End-of-Grade exams. The system is
two-tiered, with teachers in schools making ‘expected growth’ receiving a $750 bonus, and
teachers in schools making ‘high growth’ earning $1,500. The terms ‘expected’ and ‘high’

³For instance, middle school and high schools achievement gains were measured starting in 1997-98. For a
complete description of the incentive system as well as the high school criterion calculation, see Vigdor (2008).
growth are defined below.

The achievement gain threshold used in the calculation of bonus eligibility is defined as:

\[ \Delta y_{mgst} = \Delta y_{gs94} + b_1 IP_{mgt} + b_2 IRM_{mgst} \]

The \( \Delta y_{jgst} \) represents the required change in the test score for subject \( s \) for students in grade \( g \) in year \( t \) in school \( m \) compared to last year’s score on the same subject. \( \Delta y_{gs94} \) is average change in test scores for North Carolina students in 1993-94, compared to results from 1992-93.

The second and third terms on the right hand side are ‘correction factors.’ The \( IP_{mgt} \) term is the ‘Index of True Proficiency,’ and the \( IRM_{mgst} \) term is the ‘Index for the Regression to the Mean.’ The two terms are meant to adjust test score goals for shocks in performance of the school last year.

Using this criterion, for a school with \( G \) tested grades, \( 2G \) thresholds (\( \Delta y_{mgst} \)) are produced each year, which is compared to the actual average test score improvement at the school. The school scores and threshold scores are differenced, standardized\(^4\), and averaged by weight of number of students in each grade. If this average, termed ‘expected growth composite’ is greater than zero, teachers in the school receive a $750 bonus. Schools that make this criterion yet failed to test more than 98% of eligible students are exempted from the bonus. The procedure is repeated after increasing the growth threshold by 10%, to generate another average termed ‘high growth composite.’ Teachers in schools that make high growth are given an additional cash bonus of $750. Therefore, teachers in a school with exceptional test score growth scores can earn as much as $1,500 for their efforts.

The system uses End-of-Course exams to evaluate high school bonus eligibility. The general mechanism is similar to the procedure described above, but there are other elements that enter into the composite such as dropout rates.

3 Theoretical Model

A teacher’s utility function is defined as the difference between the gains to be made from the expected bonus and effort cost. I assume teachers are risk neutral in bonus receipt. Define \( e_j \in [\underline{e}, \bar{e}] \) as the effort of teacher \( j \). \( x_j \in [\underline{x}, \bar{x}] \) is the projection of teacher and

\(^4\)The difference is divided by the standard deviation of the difference across all schools in the state.
students characteristics into a single value. \(^5\) Any characteristic that positively affects student achievement increases \(x_j\). For instance, if classes for \(i\) and \(k\) are identical except the teacher in class \(i\) has more experience, \(x_i \geq x_k\). Therefore, it is natural to consider \(x\) an ability measure of both teachers and classes. \(B\) is the bonus that is paid to all teachers at the school upon qualifying under the state criterion. \(Cr\) is the state defined criterion that the school must beat in order to qualify for the bonus.

\[U_j = B \cdot Pr(e_1, e_2, ..., e_J, x_1, x_2, ..., x_J, Cr) - C(e_j)\]  

(1)

The point to emphasize here is that bonus receipt is determined at the school level. Therefore, all teachers’ efforts and all classroom characteristics contribute to the probability of bonus receipt.

Teachers maximize utility by setting their effort at \(e^*_j\) such that:

\[
\frac{\partial U_j}{\partial e^*_j} = \frac{\partial U_j}{\partial Pr} \frac{\partial Pr}{\partial e^*_j} \cdot B + \frac{\partial U_j}{\partial C} \frac{\partial C}{\partial e^*_j} = 0
\]

Because bonus is determined by school-wide achievement, it is immediately obvious that the incentive system may suffer from a free-rider (or, alternatively, incentive dilution) problem. The free-rider problem in this setting is defined as teachers providing less effort to educate her students as the number of teachers at the school increases. The econometric model will test for incentive dilution.

Another issue to consider here is whether \(e_j\) is a strategic substitute or a strategic complement. One can imagine a scenario where the coordinated effort of other teachers \(e_{-j}\) will be such that it induces teacher \(j\) to greater exertions, in which case effort is a complement. On the other hand, it is also possible that high exertion from other teachers will mean teacher \(j\) maximizes her utility by decreasing her effort, in which case effort is a substitute.

Define \(y_{jt}\) as the average classroom-level achievement in class/teacher \(j\):

\[y_{jt} = a(x_{jt}, e_{jt})\]  

(2)

Bonus is defined at the school level, where the likelihood of getting the bonus is dependent on the performance of all teachers. With \(\overline{y}_{t-1}\) defined as last year school-wide achievement, the

---

\(^5\) I normalize \(\overline{x} = 0\) and \(\overline{x} = 1\).
probability of success is defined by:

\[ P_{rt} = F \left( \sum_{j=1}^{J} \frac{a(x_{jt}, e_{jt})}{J} - y_{t-1} > Cr \right) \]  

(3)

I make the following functional form assumptions:

1. For simplicity, I assume classroom average achievement is generated as:

\[ a(x_{jt}, e_{jt}) = \exp(x_{jt}) e_{jt} \]

2. \( F(\cdot) \in [0, 1] \) is twice differentiable and \( F'(\cdot) \geq 0 \).

3. Define \( S_{-j} \in [S_-, S] \) as the average school-wide achievement excluding teacher \( j \)'s class. Then, there exists some high value of \( S, S^H \), and some low value of \( S, S^L \), such that:

\[ F(e_j | S_{-j} \geq S^H) \rightarrow 1 \]
\[ F(e_j | S_{-j} \leq S^L) \rightarrow 0 \]

for all \( e_j, x_j \).

4. (When \( J = 1 \)) There exists some high value of \( x, x^H \) and some low value of \( x, x^L \), such that:

\[ \frac{\partial F(\cdot | x_j \geq x^H)}{\partial e} \approx 0 \]
\[ \frac{\partial F(\cdot | x_j \leq x^L)}{\partial e} \approx 0 \]

5. The effort cost function, \( C(e) \) is twice differentiable, with \( C'(\cdot) > 0 \) and \( C''(\cdot) \geq 0 \).

**Proposition 1** Given Equations (1) - (3), Assumptions (1) - (5), and \( \{x, B, Cr, J\} \) there exists an interior pure strategy Nash equilibrium in effort, \( \{e_1, e_2, ..., e_J\} \).

Assumption 3 is a statement about the limitation of effort within the school. That is, one teacher cannot unilaterally determine the bonus receipt of the entire school. Therefore, if all other teachers shirk, the best response of teacher \( j \) is also to shirk. On the other hand, if all teachers are giving maximal effort such that the bonus is assured, the best response of teacher
$j$ is again, to shirk. It is exactly when all other teachers are putting forth some amount of effort and the bonus receipt is in doubt, that teacher $j$ is also induced to give some positive amount of effort.

Assumption 4 is a statement about the limitations of effort within the classroom. It defines the limits of the effect of effort. That is, if teacher/class ability is very low (or very high), the teacher, through the application of effort, cannot substantially increase her probability of receiving the bonus. When ability is very high, the probability of students meeting the state criterion is 1 already, and additional effort from the teacher does not increase the probability. When ability is very low, additional effort falls on ‘deaf ears’ (or the teacher is so inept that her efforts are misdirected), and the probability of receiving the bonus remains at or near zero.

As shown in the existence proof, when $e_j$ and $S_{-j}$ are increasing together, effort is a strategic complement, when $e_j$ declines in $S_{-j}$, effort is a strategic substitute. Whether effort is a complement or substitute is determined in the model by the shape of the $F'(\cdot)$.

As alluded to earlier, the incentive system may suffer from a free-rider problem. I show that there can exist a free-rider problem.

**Proposition 2** Assuming identical teachers and classes, $x^L < x < x^H$, and $S^L < S < S^H$, a free rider problem may exist.

In the model, the free-rider effect can arise from two separate channels. The first, direct channel comes from the increase in the number of teachers from $J$ to $J + 1$.\(^6\) As $J$ increases, a teacher’s effort is distributed over a larger population. While the size of the potential ‘pie’ increases as the number of teachers increases, the ‘piece’ going to each teacher stays constant. Since the teacher’s effort is being distributed over a larger number of agents while her share remains constant, she will be induced to lower her level of effort. In this sense, free-rider problems always exists.

There is a second, indirect channel to consider. An increase in the number of teachers in the model necessarily implies an additional class.\(^7\) This means that $x_{J+1}$ is included in the probability of receiving the bonus, which can change the distribution of ability of classrooms.

---

\(^6\)Note that in this context, an increase in $J$ is a discrete increase. $J + \epsilon$ is a meaningless concept here.

\(^7\)I assume away the possibility of $J$ increasing, the number of classes decreasing, and average class size decreasing.
in the school. In proposition 2, I assumed that the distribution of abilities does not change (due to homogeneous teachers and classes), to negate the indirect channel. However, it is easy to imagine a scenario where the new teacher/class ability is far away enough from the school average to have a large effect on effort exertion across the school.

From the first order condition:

\[
\frac{1}{J} BF'(y_m) \exp(x_j) = C'(e_j)
\]

I will assume that the cost function is quadratic for the empirical section and the probability distribution is standard normal. Taking natural logs, the FOC becomes:

\[
\ln e^*_j = \gamma + x_j - \ln J + \ln \phi(y_m(x_1, x_2, ..., x_J))
\]

where \(\gamma\) is a normalizing term. The two effects from the bonus being paid out by school: the direct free-rider effect arising from an increase in \(J\) and indirect effect arising from the change of distribution of ability, are separable. This will greatly simplify the econometric specification.

Having found that school incentive policy may suffer from a free-rider problem, a simple solution would be to go from a school-wide incentive to a classroom-level (non-cooperative) incentive where teachers are judged by her students’ performance. The first order condition shows that the \(-\ln J\) would equal zero when \(J = 1\), thus eliminating the direct free-rider effect altogether. However, I show below that moving to this non-cooperative criterion will not necessarily increase effort exertion of teachers.

**Proposition 3** It is possible for effort to decline when the regime changes from cooperative to non-cooperative.

The key is that while decreasing \(J\) to 1 will increase effort by decreasing the free-rider problem, moving from school-wide average to classroom average will move each teacher to a different point on \(F(\cdot)\). Whereas a teacher may have been at a point on the distribution where marginal effort can make a lot of difference in the cooperative criterion, the non-cooperative criterion may place her at a point where marginal effort exertion makes little difference in change in the probability of bonus receipt.

\[8\]I ignore the political resistance that may arise.
This property is best demonstrated by looking at a school with two teachers. Assume teacher 1 has $x_1 > x^H$ and teacher 2 has $x_2 < x^L$. These teachers’ optimization solutions are demonstrated in Figure 1 and Figure 2.

The thick lines represent indifference curves of teachers, and utility increases up and to the left. In an individual criterion regime, as the probability of bonus receipt is flat for both teachers (Assumption 3), their optimum solution is at the corner: $e_1 = e_2 = 0$. Now, assume that bonus is awarded for joint performance. In this case, the teachers’ optimum solutions move to interior points, and $e_i > 0$ for $i = 1, 2$, as marginal effort exertion for both teachers will now increase the probability of bonus receipt.

Surprisingly, the theory results here indicate that a collaborative regime that experiences some free-rider problems may be more desirable than a non-cooperative criterion environment depending on the distribution of ability. If schools are composed mostly of homogeneous teachers and classes, the free-rider problem may result in lower overall effort exertion. If the bonus system depends on individual achievement, and if the state criterion is set correctly such that the teachers are induced to work harder to attain the bonus, student achievement may increase. On the other hand, if schools are roughly composed of very high and very low ability teachers and classes, a cooperative incentive policy may actually induce greater effort and increased achievement. The policy simulation section explores this theoretical implication more fully.

---

9I assume that the higher-ability teacher’s marginal effort application is more effective compared to the lower-ability teacher, but the solution holds if I assume the opposite.

10A simple example illustrates this point: Imagine two 100 meter runners in the Olympics. One is the world record holder, and the other has the slowest time amongst the competitors. In individual competition, both runners exert a low amount of effort. The fast runner can win with relative ease, while the slow runner merely walks to the finish. When these two runners are part of a relay team, both will exert higher effort. The fast runner must increase his effort to increase chances for victory. The slow runner now has a chance for victory because of the fast runner, and must run faster than his individual time to contribute to the joint victory.
4 Data

I use an administrative dataset for the North Carolina public school system from the academic years 1999-00 to 2003-04. The dataset contains information on all public schools, students, teachers, and administrators in the state of North Carolina. Since the data is collected annually and individuals can be matched across years, a relatively complete longitudinal picture of the entire public school system in North Carolina emerges, detailing students’ academic trajectories, peer interactions, and exposure to teachers.

There are two unique features of the data that I take advantage of to identify the effect of teacher effort. The first is that each student record is linked to a teacher identification number. This permits the identification of a complete classroom, with information on the student, teacher, and peers, provided that student instruction is confined primarily to the self-contained classroom. While most students in middle schools (grades 6 through 8) and high schools (grades 9 through 12) change teachers and peers each period, elementary school students are tied to a single classroom, where they are exposed to the same peers and teacher throughout the school day. Therefore, any effect of effort from the elementary school teacher will be isolated to her classroom.

The data set tracks student performance from year-to-year as long as they remain in the North Carolina public school system. Because of the need for two years’ worth of performance for each student to judge whether the student has improved, all students with only a single year record are dropped. Background characteristics information such as sex, race, age, and parent’s education level are collected. I divide race into minority (black and Hispanic) and white (all others). I divide parent’s education level into those who have high school education or less, and those who have above high school education.

The second unique feature of the dataset is that it collects absence-taking behavior data of teachers. Because an elementary school student is exposed to one teacher, if that teacher takes a day of absence, the effect will be isolated to her students only. Teacher absences in North Carolina are largely categorized into: sick leave, personal leave, and annual (vacation)

---

11The data, which is collected by NCDPI, was made available by NCERDC (North Carolina Education Research Data Center: www.pubpol.duke.edu/centers/child/nceddatacenter.html) at the Center for Child and Family Policy. While student and teacher level data are confidential, aggregate data and summary statistics are publicly available at the NCDPI web site (www.ncpublicschools.org/reportstats.html).
leave. I use the sum of sick leave and personal leave as the measure of teacher absence in an academic year. The data shows that most of the annual leave days coincide with school vacation days.\textsuperscript{12}

Because sick and personal leaves are unplanned-for and take place during the school year, they have the most effect on student learning. For the study period, teachers took an average of 9.6 days of absences in an academic year of roughly 180 days. I restrict the sample of teachers to those who take less than or equal to 30 days of absences in the academic year. There are significant numbers of teachers who take more than 30 days off (about 5\% of the original sample). This is due to the fact that experienced teachers can accumulate sick leave (at a flat rate of one a month, indefinitely and without upper limit), and may choose to take very long periods of time off. I exclude these teachers from analysis.\textsuperscript{13} The resulting sample absence rate of 5.3\% is in line with other studies of teacher absences.\textsuperscript{14} Table 1 summarizes student, teacher, class, and school characteristics and absence-taking behavior.

5 Econometric Model

The econometric model will estimate three equations, following the theoretical model. School-level expected bonus is estimated using student achievement and the accountability rules defined in section 2. The teacher’s effort will be estimated using the observable measure, teacher absence decisions. I show that teacher absence can be separated into effort and other unrelated shocks, and use the incentive legislation to isolate the effort component. This reduces noise arising from uncorrelated shocks, making teacher absence a good signal for effort. Finally, student achievement is estimated using signaled effort from the second equation along with other traditional student, peer, teacher demographic characteristics. The system is solved iteratively until convergence.

\textsuperscript{12}About 60\% of annual leaves are concentrated in December, June, July, and August. Another 20\% is taken in November, which indicates Thanksgiving break.

\textsuperscript{13}Including teachers with high level of absence does not qualitatively change results, but makes absence much more dependent on a teacher’s experience status. Table 2, in the results section, shows that experienced teachers on average take almost 4.5 more absences even if I restrict the sample to teachers with less than 30 days absence.

\textsuperscript{14}See Ehrenberg et. al. (1991).
5.1 School-level Bonus

While the expected bonus would usually be estimated as a probit or logit from data on bonus receipt, last year’s test scores, and this year’s predicted scores times the bonus amount, I can use the state defined criterion as specified in the NC accountability system description. Because the differences between school scores and state criterion are standardized to N(0,1), let \( \hat{y}_{st} \) represent the predicted school performance this year, and let \( y_{s(t-1)} \) be school performance last year. Expected bonus is simple to calculate:

\[
E(B) = \{ \Phi(\hat{y}_{st} - y_{s(t-1)} \geq Cr_1) + \Phi(\hat{y}_{st} - y_{s(t-1)} \geq Cr_2) \} \cdot B
\]

where \( Cr_1 \) and \( Cr_2 \) are expected and high growth composite threshold values, respectively, and \( B \) is the bonus amount.\(^{15}\) Predicted scores for this year’s tests are defined in subsection (5.3).

5.2 Teacher Absence Decision

If effort were readily measurable, I could estimate the effect of incentives by running the following regression:

\[
e_{jst} = \alpha_0 + X_{jst} \alpha_1 + \alpha_2 I_{jst} + \epsilon_{jst}
\]

where \( e_{jst} \) is the teacher’s level of effort, \( X_{jst} \) are observable characteristics of teacher \( j \) and her school \( s \) that are regularly used in the empirical economics of education literature, \( I_{jst} \) is the measures of incentive strength (bonus receipt probability and incentive dilution), and \( \epsilon_{jst} \) is the idiosyncratic error.

While effort is not directly measurable, teacher absence taking behavior serves as a noisy signal for effort. Teacher absence is noisy because it encompasses effort as well as unrelated shocks. This subsection demonstrates how to decrease the noise and distill the signal for effort from absence.

I hypothesize that teacher absence, \( A_{jst} \) is determined by three different components, as defined below:

\[
A_{jst} = g(X_{jst}, e_{jst}, \eta_{jst})
\]

\(^{15}\)Note that in the estimation, following the first order condition, I use the pdf values \( \phi(\hat{y}_{st} - y_{s(t-1)} \geq Cr_1) + \phi(\hat{y}_{st} - y_{s(t-1)} \geq Cr_2) \).
\( \eta_{jst} \) represent factors that affect absence that are unrelated to \( e_{jst} \), such as unforeseen bad/good health outcomes, unexpectedly bad/good weather patterns, and other shocks that affect the teacher’s absence outcome. For simplicity, if \( e \) were directly observable, I assume I can estimate the absence decision using an OLS framework.

I assume that for two teachers \( i \) and \( j \), \( A_i \geq A_j \) if and only if \( e_i \leq e_j \), all else equal. Because the absence variable is a count of days of absence in the entire school year, it is appropriate to think of effort as the aggregate effort provided by the teacher throughout the entire school year. The condition assumes away situations where a teacher may redistribute effort for different time periods. For instance, a teacher cannot take an absence, ‘save up’ her effort for that day, and saturate her students close to exam time.\(^{16} \)

Since absence \((A_{jst})\) is correlated with effort \((e_{jst})\), and incentives \((I_{jst})\) are correlated with effort, the projection of \( I \) on \( A \) results in an unscaled measure of effort.\(^{17} \) Using the data available and the first order condition directly from the theory, I model the absence decision as follows:\(^{18} \):

\[
\ln A_{jst} = X_{jst}\alpha_1 + \alpha_2 \ln J_{st} + \alpha_3 \ln (\phi(\hat{y}_{jst}, Cr_1) + \phi(\hat{y}_{jst}, Cr_2)) + \eta_{jst}
\]  

(5)

The measure of incentive strength \((I_{jst})\) to be used are \( J_{st} \), the number of teachers at the school, and \( \phi(\hat{y}_{jst}, Cr_1) + \phi(\hat{y}_{jst}, Cr_2) \), the position of the school with respect to the probability of bonus receipt. The \( J \) term estimates the direct effect of free-ridership, as discussed in the theory. The effect of the \( \phi \) term is best explained by Figure 3. The \( \phi \) term measures the incentive effects of being at different points in the distribution of bonus receipt probability. The shape of the log of the pdf is strictly concave, with low effort tied to very low and very high achievement. The peak of effort exertion is tied to middle level achievement, when the bonus receipt is in doubt and very dependent on the effort exertion of teachers. Since absence is negatively associated with effort, I expect \( \alpha_3 < 0 \). The shape of the \( \phi \) variable is important, as it reflects the non-linear effect of the probability of receiving the bonus. As bonus becomes

\(^{16}\)See appendix for a formalization of this condition.

\(^{17}\)This relationship relies on the assumption that \( \eta \) is uncorrelated with \( I \). That is, higher or lower incentive strength is not likely to cause good or bad health/weather shocks.

\(^{18}\)Recall that the first order condition from the theory is:\(^{19} \):

\[
\ln e = \gamma + x - \ln J + \ln \phi(\cdot)
\]
very easy or very difficult to attain, the utility maximizing response of teachers is to decrease their effort expenditure. The non-linear effect will be important for policy analysis. \( \eta_{jst} \) is the idiosyncratic error term that represents the uncorrelated health and weather shocks.

By estimating absence decision in this way, I decrease the noise in teacher absence and distill it into a useful signal for effort. The change in ‘predicted absence’ driven purely by the incentives in the bonus program is an indicator of change in effort, because absence reduction \textit{per se} is not rewarded as part of the incentive system.\(^{21}\)

### 5.3 Student Achievement

The student achievement equation does not follow from any specific utility maximization solution for students. I assume that there exists some production function for education where the inputs are student, peer, and teacher characteristics. Of primary interest will be the effort input of teachers. The achievement function for student \( i \) is specified as follows:

\[
y_{ist} = Z_{ist} \beta_1 + \beta_2 y_{ist(t-1)} + \beta_3 \ln(\hat{A}_{jst}) + \beta_4 \ln(\hat{A}_{jst}) \cdot I[\text{minority}] + \nu_{ist} \tag{6}
\]

\( Z_{ist} \) is a vector of student, peer, and teacher demographic characteristics. \( \hat{A}_{jst} \) is the projected measure of (negative) effort from equation (1). \( \nu_{ist} \) is the idiosyncratic error. The reason for the inclusion of \( \hat{A}_{jst} \) is to estimate the effect of year-long teacher effort on student achievement.\(^ {22} \)

\(^{20}\)An alternative specification may shed light on the role of \( \phi \). To check to see if the shape of the pdf is consistent with the I replace \( \phi \) with a monotone measure \( E(B) \), the expected bonus:

\[
A_{jst} = X_{jst} \alpha_1 + \alpha_2 J_{st} + \alpha_3 E(B)_{st} + \alpha_4 E(B)_{st}^2 + \eta_{jst}
\]

Note here that \( E(B) \) also has a squared term. The squared term of expected bonus is included to capture non-linear effects of the probability of receiving the bonus. If the shape of the pdf of the distribution is correct, I expect \( \alpha_3 > 0 \) and \( \alpha_4 < 0 \). Estimation of this alternative specification is presented in Table 6, and results support the functional form assumption.

\(^{21}\)That absence is not part of the incentive system is important for its role of serving as a signal for effort. If absence behavior were part of the incentive system and teachers were aware of this, depending on the state formula, teachers could choose to simultaneously reduce effort and absence, while ensuring the same or higher pay out. This would render absence useless as a signal.

\(^{22}\)It is possible to argue that \( J \) should be included directly in the achievement equation, as a measure of individualized attention to students. Note further that \( J \) as a measure of individualized attention, and \( J \) as a measure of free-ridership, must pull achievement in opposite directions. Therefore, to test whether \( J \) has a direct \textit{positive} effect on achievement, I run an alternative specification using \( J \cdot I[\text{no bonus effect}] \) as an exclusion
I make two concessions to expediency in estimating the achievement function. The first is that the achievement function defined in the theoretical model implies effort is completely separable upon taking logs.

\[ y_{ist} = Z_{ist} \beta_1 + \beta_2 \ln(A_{jst}) + \nu_{ist} \]

Because the difference in sensitivity of whites and minorities is an important component of the model, I include an interaction term of log of absence and an indicator variable for minority students. This creates two effectiveness measures of effort, one for white students and one for minority students. As I show in the results and policy simulation sections, the difference in effort effectiveness across races is important for the teacher utility optimization solution as well as the effectiveness of suggested policy changes.

The second concession is the inclusion of lagged achievement. Lagged achievement does not appear in the theoretical model, and the inclusion of \( y_{is(t-1)} \) creates endogeneity problems. However, last year’s test scores are included to reduce the possibility of omitted variables (such as developmental factors) introducing bias. The consensus in the literature is that benefit from using the value-added approach far outweigh problems caused by the endogeneity of the variable (Rivkin, Hanushek, and Kain (2005)). Furthermore, the North Carolina testing system is graded on a developmental scale, making it natural to compare scores from subsequent years.

I could include \( y_{is(t-1)} \) in the dependent variable and estimate a “gains” equation. However, this assumes that there is zero educational decay from one year to the next (after students take summer break). The functional form in this study assumes a constant rate of decay across all students and grades. An alternate form of controlling for developmental factors (that a student is exposed to at school) would be include a vector of the student’s previous teacher and peer exposure and estimate heterogeneous rates of education decay (Rothstein (2008)).

I restriction and find that there was no statistically significant direct effect of \( J \) on \( y \). Here, \( I[\text{no bonus effect}] \) is an indicator variable which equals 1 when the probability of obtaining the bonus is near 100% or 0% and 0 otherwise. That is, either the bonus is assured or out of reach, meaning that the effect of \( J \) should only come through the direct channel. See Table 8.
6 Results

Having specified the estimation strategy, in this section, I present the results. Before the parameter estimates are considered, it is natural to wonder if the procedure to distill the signal is required at all. A single equation OLS or fixed-effects model that uses raw absence data may suitably summarize the relationship between teacher effort and achievement.

To motivate the need for signal distillation, I start by showing OLS estimates without isolating the effort component in the first step. In essence, I estimate equation (6) using raw absence data. The results are shown in the first column of Table 3. The parameter estimate on absence is similar to results from Clotfelter, Ladd, and Vigdor (2007), and shows that teacher absence has a small, negative impact on student achievement. The results indicate that on average, students do not suffer much from teacher absence. Enacting some incentive policy that would cut the average absence rate in half, from approximately 10 days to 5 days, would increase average achievement by roughly 0.5% of a standard deviation. In contrast, paying for teacher certification would be approximately 6 times as effective.

The reason for the smaller than expected OLS estimate is because an absence by the teacher does not result in zero education production. A substitute teacher will be assigned and students will still attend class. Imagine two identical teachers teaching identical classes. If one of the teachers takes one additional day off compared to the other teacher (with a substitute teacher filling in), student performances should not be significantly different across the two classes.

The structural model estimates the impact of ‘predicted absence’ due to incentive effects at more than 30 times greater than the OLS parameter estimates. This large difference can be explained by extending the example above. Assume the two teachers now differ on effort. One teacher is more motivated than the other teacher. The unmotivated teacher does not prepare for instruction, pays poor attention to students, and generally does not care about the educational outcome of students. Further assume that these teachers are in charge of identical classes. All else equal, student achievement in the unmotivated teacher’s class would

\[\text{Clotfelter, Ladd, and Vigdor (2007) shows that fixed-effects estimates yield smaller but still significant estimates of the effect of absence. The reading scores are adjusted to represent percentage change of one standard deviation of scores.}\]

\[\text{Including fixed-effect parameters at the student level yielded no qualitative differences.}\]
be lower, and the teacher would take more absences. Now, assume that the enthusiastic teacher has a bad health shock and takes as many days off as the other teacher. While students in both class are exposed to their teacher the same number of days, the enthusiastic teacher will give superior instruction to her kids throughout the entire academic year. The OLS procedure in this case treats the teachers identically (due to the same number of absences), and predicts the same achievement. The structural model predicts higher effort and achievement in the motivated teacher’s class, due to the incentives from the accountability system.

Since the change in absence arising purely from the incentives is small relative to overall fluctuations in absence (due to other uncorrelated shocks), it is unsurprising that the effort parameter estimate is so much larger. Because of this difference, observed absence and predicted absence are fundamentally different measures. Each incremental increase in predicted absence is reflected in the quality of instruction for the entire academic year. Therefore, the parameter estimates for $J$ and $\phi$ should be regarded as the effect of incentives on effort, where differences in levels of predicted absence represent differences in average effort level provided by the teacher throughout the year.

The first-stage estimates show that a teacher exert lower levels of effort when the number of colleagues in the school increases. A higher number of teachers at a school implies lower effort for teachers, ceteris paribus. This points to the existence of incentive dilution (or free-ridership) in school-level incentives.\footnote{A possible concern here is that the parameter on the number of teachers is not measuring free-ridership arising from the bonus regime, but a pre-existing condition. I run a regression that measures the sensitivity of absence on the number of teachers in the school for pre- and post-incentive system installation. The parameter estimate on the number of teachers pre-incentive is statistically insignificant, while it is positive and significant for parameters from post-incentive installation. See Table 7.}

The sign of the parameter on the $\phi$ term shows that very low and very high probabilities of bonus receipt are associated with low effort exertion. Translating the $\phi$ term into expected bonus, at low levels of expected bonus, teachers exert low effort, with effort increasing as expected bonus (probability of bonus receipt) increases. The peak level of effort exertion is achieved when expected bonus is at approximately $873, at this point, the bonus becomes a ‘sure-thing,’ and effort declines again toward low levels as expected bonus increases. Therefore, if the goal of the incentive policy is to get teachers to exert the maximum amount of effort, the
threshold value for attaining the bonus must be set such that qualifying for the cash bonus is neither too easy nor too difficult.

While the estimates for effort signal show how teachers respond to incentives, the estimates for effect on achievement show how students respond to these motivated teachers. Increasing expected bonus amount from $400 to $800 will increase student achievement for white students by about 8.7% of a standard deviation on their end of grade reading exam, and about 7.9% of a standard deviation for minority students. White students are more sensitive to teacher effort by about 9% compared to minority students.

This impact of teacher effort on student achievement, the impact of incentive criterion on expected bonuses, and the impact of expected bonuses on teacher effort are summarized in Figure 4. The top-left quadrant portrays the effect of predicted absence on student achievement. The top-right quadrant shows the non-linear impact of expected bonus on a teacher’s effort. The bottom-right quadrant shows the relationship between the severity of the bonus criterion and the expected bonus. Finally, the bottom-left quadrant of the graph shows the non-linear effect of incentive criterion severity on student achievement which arises directly from the teacher’s non-linear response to expected bonus. This shows that policy makers must take care in setting the right criteria to motivate the teachers. A too lax or too strict a criterion will actually drive down student achievement, and potentially waste money.

Inefficient policy design which wastes funds is also illustrated in the figure. Assume that the state target for student achievement is at point A. There will be two levels of criterion severity that accomplish this: points E and D. Point E, the tougher standard, is associated with pay-out point B, and point D, the easier standard, is associated with pay-out point C. The two policies, ABEF and ACDG, yield the same student performance. However, policy ACDG pays teachers the horizontal distance between B and C extra for no additional achievement gains. It is possible that the increase in expected pay can act like a pay increase, and change the composition of the teacher population. This could lead to higher quality teachers and higher educational achievement. However, estimation of this effect is beyond the scope of the model.

While the ability to measure teacher effort and its effect on student achievement is interesting in its own right, the more important question is how to design policy to effectively induce effort. The next section presents three possible policy changes. The first experiment changes
the criterion from a school-wide performance to classroom performance measure, attempting
to eliminate the incentive dilution problem. The second experiment combines the NC system
and the NCLB system, by explicitly tracking the performance of minority students. The third
experiment looks at making the current standard stronger or weaker, keeping all other details
of the bonus system identical.

7 Policy Simulations

In the previous section, I showed that the incentive system suffers from free-rider problems,
and teachers respond in a non-linear fashion to cash incentives. Therefore, setting arbitrary
criteria or cash awards and hoping for increased performance may be inefficient or even coun-
terproductive. I present three possible policy alterations to the current system to examine
their possible effects on teachers and students.

In particular, I focus on disparate achievement growth of white and minority students.
The primary driving force behind the disparity is the fact that the effect of teacher effort is
more pronounced for white students compared to their minority counterparts. White students
are more sensitive to teacher effort, by about 9%. Combined with their absolute lower level
of achievement, minority students have a lower and shallower trajectory of academic growth
for levels of teacher effort applied, compared to white students. The differences in marginal
impact of effort implies that effort equally applied to whites and minority students will affect
their scores differently, thus affect the teacher’s expected bonus differently.

7.1 School-wide to Individual Classroom Standard

As discussed in the theory section, schools under the North Carolina bonus system may suffer
from a free-rider problem. With individual bonuses depending on the effort of all teachers,
it is not surprising that more absences are predicted as the number of teachers in the school
(controlling for student population) increases.

One may argue that teachers should be evaluated on individual performance so that free-
riding is no longer an issue. I change the state formula for bonus receipt from school average
scores to class average scores to generate new expected bonuses for each teacher. Further, I
fix the number of teachers in the school to one, and estimate effort. The results are shown
in the second column of Table 4. The results from the simulations show that expected bonus on average increases slightly. Yet at the same time, average predicted effort decreases by about 4% and variance of expected bonus (and by extension effort) increases. The decrease in average effort is the opposite of the intended effect of going to a non-cooperative criterion. As a result of the decrease in effort, state average reading score declines by a substantial margin. However, a slight majority of teachers give more effort in the new regime, leading to a slight majority of students benefiting from the policy.

These seemingly contradictory results can be explained by examining teacher incentives. With the regime switch from school-wide to classroom, teachers would be split sharply into those who can achieve the bonus standard, and those who cannot. Going from a collaborative effort to an individual effort will starkly separate low performers from high performers. A teacher who determines that she has a low chance of success in the classroom regime will have little incentive to exert effort, while a teacher with a fair chance of success will only increase her effort until marginal benefit matches marginal effort cost. In addition, teachers with very high achieving students will decrease effort, since the probability of bonus receipt is no longer dragged downward by other lower achieving classes in the same school. With the right and left tail of the expected bonus distribution pushing effort downward, average effort decreases.

The histograms in Figures 5, 6, and 7 demonstrate the effect of switching from a school-wide to a classroom criterion. Figure 5 shows that going to classroom performance increases the weight of the tails of the expected bonus distribution. The right tail represents classrooms which were outperforming the rest of the school, yet were being dragged down as part of the school-wide criterion. The left tail represents classrooms which were under performing compared to the rest of the school, yet were buoyed by the being part of the school-wide criterion. Figure 6 shows that as the bonus outcome changes, teachers adjust their efforts accordingly. While there is little change in high effort part of the distribution, the percentage

---

26 Expected bonus increases by about $11.63. Since there are roughly 47,000 teacher/year observations over the five-year sample period, and assuming that there is an equal number of kindergarten, first, and second grade teachers in the qualifying schools, gross bonus pay out would increase by roughly $1 million each year. Considering the state budget for bonus pay out is approximately $90 million dollars, this is not a substantial increase.

27 In this context, ‘low’ performer does not necessarily indicate poor teachers, but teachers with low probability of success. This could be the result of the composition of her class, for example.
of teachers taking higher levels of absences increases. This low effort teachers increase is directly attributable to the increase in the left and right tail of the first histogram. Teachers of classes with highly under/over performing classes are no longer pushed toward the middle of the distribution by other classes in the school, inducing these teachers into lower effort exertion. The result of the increase in low teacher effort is reflected in the increase in low performance and the slight decrease in high performance as illustrated in Figure 7.

An interesting fact is that despite the decline in average scores, a slight majority of students benefit from the policy. This is because while the elimination of free-riding behavior benefits a majority of students, the students who are taught by teachers who ‘give-up’ under the new regime are hit particularly hard, as teachers lower effort exertion.

While the simulation results show that a slight majority of students benefit, the statewide average declines with the racial achievement gap widening. This indicates that the majority of those experiencing declines in scores are minorities, and the beneficiaries of the policy would be white students. Teachers with minority-heavy classes would be induced to pull back effort by large margins, resulting in performance degradation. This shows that policies to eliminate the free-rider problem would actually be regressive, hurting students who require the most help.

In addition to these mixed results, several problems arise with switching to the proposed standard. First, there could be large, negative general equilibrium effects. For instance, teachers may compete to avoid students who are least likely to make year-over-year improvements. With experienced teachers having seniority, we may expect the students who require the most help ending up with the least experienced teachers. Another possibility is that teachers who are perpetually stuck with under-achieving students may seek to transfer or exit the profession altogether.

Second, as the sample size of the signal decreases from school scores to class scores, the variance of the signal also increases. Depending on the degree of risk-aversion for teachers, the increased variance may lead to lower levels of effort.

Third, teachers in non-tested grades would need to be compensated in some alternate manner. Currently, kindergarten, first, and second grades students are not administered an end-of-grade exam. Elementary school teachers in charge of these students are paid the bonus according to the school-wide performance. With a change in regime, the state would need
to administer tests to all grades. In evaluating a move to individual criterion, these negative factors must be considered carefully, along with the positive results from the elimination of free-ridership.

### 7.2 Race Specific Requirements

Although NCLB has many critics, one of the components of the policy that is generally regarded in a positive light is the emphasis on achievement goals for historically under achieving students. NCLB divides the school population into racial and socioeconomic subgroups and evaluates the performance of each subgroup separately. A school must ensure that every subgroup passes the NCLB criterion, or the entire school fails.

No such group-specific achievement targets are specified in the North Carolina system. I simulate the teachers’ responses if the bonuses were to depend purely on the achievement of minority students. Since white students are more sensitive to teacher effort, if minority-only standard induces overall teacher effort increase, the benefits will be shared by all students, regardless of race. The result is illustrated in Figure 8.

The response of teachers in the minority-only standard is dominated by the status quo standard at higher levels of expected bonus. That is, at high levels of expected bonus level, teachers would exert lower levels of effort in a minority-only standard. Because minority students are less responsive to teacher effort, teachers have less incentive to raise the possibility of attaining the bonus through higher effort.

However, at lower levels of expected bonus, the difference between the two regimes are relatively small, and in the region between $200 and $600, the minority-only regime actually induces slightly higher effort than the status quo.\(^{28}\)

The small difference is attributable to the fact that at lower levels of expected bonus, schools have a much higher proportion of minority students. Therefore, the minority-only standard and status quo are measuring essentially the same schools at lower levels of expected bonus. As expected bonus increases, schools have higher proportions of white students. The differences between the regimes become stark at this higher level of expected bonus receipt. Teachers in the status quo regime can be induced to increase their effort by about 5% compared

\(^{28}\)The largest difference between the two regimes occur at $450 and teachers exert about 1.5% more effort in the minority-only regime compared to the status quo.
to those in the minority-only standard at their respective highest levels of effort. This is because it is very difficult for schools operating under the minority-only regime to attain the high growth standard to qualify for the second bonus. The results are summarized in the third column of Table 4.

The results show that on average, teachers exert lower effort as the probability of attaining the bonus declines in the new regime. Expected bonus declines per teacher by about $40, which roughly translates to about $4 million dollars in savings per year. Only one in ten teachers increase their effort level, and one in ten students improve achievement. Interestingly, students who experience achievement gains are concentrated in the minority students and those who have parents with no college education. This is because there is a subset of schools that is just at the point of receiving the bonus, and the change in regime slightly decreases the chance for bonus receipt. The student population at these schools would have to be skewed toward minorities, as a predominantly white school would experience a larger decline in probability of bonus receipt. In response, teachers in this particular subset increase their efforts, delivering more education to the students, thus raising their achievement. As a result, the racial achievement gap can be seen to close slightly.

One caveat to this calculation is that I rely on an underlying assumption about how teachers divide their attention among the students in the classroom. If we believe that teachers give equal attention to all students in her class (as I assume), evaluating her bonus receipt qualifications using only the test results of her minority students may in fact induce her to give forth less effort and lower the scores of all students. While the racial achievement gap will decrease, since white students are more sensitive to absences, this result is achieved by lowering both white and minority achievement.

However, if a teacher’s class is almost completely white with one minority student, and the teacher is told that her bonus depends only upon the performance of the minority student, it would be naïve to assume that she would spread her efforts evenly across all students. The best response function shows that the teacher would decrease effort, but she may redistribute her new level of effort to focus on the minority student. In this case, the minority student’s score will rise, to the detriment of his white peers. While this will also close the racial achievement gap, the actual calculation of the magnitude of the change is beyond the scope of this paper, since I do not specify a model for redistribution of effort within the class.
In either case, the teacher will decrease her average level of effort, leading to an average decline in student achievement. Whether the teacher redistributes her effort to favor minority students or not, the best case scenario for closing the achievement gap in a minority-only regime is by raising minority achievement while simultaneously lowering white achievement.\textsuperscript{29}

### 7.3 Tougher/Easier Bonus Requirements

If students are not performing as well as hoped with the incentive regime and if we believe that teacher effort makes a difference in raising test scores, it may seem logical to induce higher effort by raising the standard at which bonuses are paid.\textsuperscript{30} Furthermore, this may have the added benefit of lowering the overall bonus pay out to teachers from the state, allowing the money to be spent elsewhere. For instance, the surplus may be used to raise teacher base salaries. However, there are two potential problems with this tactic.

First, merely strengthening the criterion, thereby lowering the probability of qualifying for the bonus, will not necessarily encourage higher effort. As discussed in the results section, if the criterion is already too strict, further toughening the standard will only induce lower effort as teachers resign themselves to losing the bonus, translating into yet lower scores. Toughening standards will be effective when the current criterion is too lax, allowing teachers to coast while qualifying for the bonus.

Second, because each school has a different probability of making the bonus requirement, strengthening the statewide standard will have differing effects for different schools. For instance, teachers in a high growth school that had little trouble making the bonus requirement under the lax criterion may now be forced to work harder, producing even higher growth rates. However, teachers in low growth schools may actually put forth less effort, as the marginal gains from bonus probability shrinks below the marginal cost of current effort.

If one of the goals of the policy is to close the gaps between privileged and underprivileged students, a stricter criterion makes sense only if high growth schools are actually schools that educate traditionally underprivileged students. In order to evaluate the effect of a stricter

\textsuperscript{29}Since the average achievement declines, I would need to specify a social welfare function that weights the welfare of minority students heavier than the welfare of white students to justify this policy as ‘welfare-improving’ or socially desirable.

\textsuperscript{30}Raising the bonus amount will unequivocally raise the effort level of all teachers. However, as this ignores a budget constraint problem, I choose not to explore the policy ramifications.
policy, I simulate a 10% increase in the ‘expected’ and ‘high growth’ criterion, and evaluate the change in predicted absence of teachers, and subsequent performance change in students. The results for the tougher criterion are presented in the second column of Table 5.

After the criterion is strengthened, the average teacher effort in the state decreases by about 2.3%. As a result, average reading performance in the state declines. This indicates that on average, the criterion was already too strict. Furthermore, the losses are concentrated in minority students with parents who have a high school degree or less, resulting in an increase in the racial achievement gap. Due to the increase in criterion strength, the average pay out per teacher declines by about $57.70 resulting in about five million dollars in savings per year. However, the decrease in effort is not uniform across the state. A small fraction of teachers actually exerts more effort. Those teachers who increase effort are more likely to teach in schools that have a higher proportion of students who are white and have highly-educated parents.

The third through sixth columns of Table 5 show the results of making the standard more lax. Easier to achieve standards induce higher levels of effort among teachers, resulting in higher student achievement. This comes at the cost of substantial increases in aggregate pay outs to teachers, possibly resulting in reduction of base pay for teachers.31

Most encouraging is the fact that the achievement gains are more concentrated in the minority student population, reducing the racial achievement gap. However, as the last two columns show, once the criterion becomes too easy to satisfy, all teachers begin to cut back effort. These results point to the need to more carefully determine how to set the strength of the criterion to maximize teacher effort.

8 Conclusion

This study sought to measure one of the primary determinants of education production, teacher effort, and examine the effectiveness of accountability legislation in altering teacher behavior. In particular, the effect of effort on overall student achievement and the racial

31Unlike previously considered policies, this would be a much heavier burden for the state to bear. For instance, making the criterion 20% easier to meet would raise the expected bonus by about $110. This would translate to more than a $10 million dollars per year increase in bonus pay outs to teachers.
achievement gap was examined. Using the absence taking behavior of teachers as a signal for effort and the incentive system in North Carolina, I find that teachers respond to cash incentives, and effort makes a substantial difference in student achievement. I also find that white students are more sensitive to teacher effort than their minority counterparts, and teachers respond to this difference in sensitivity by altering their effort exertion.

I performed three policy simulations to gauge the possible effects of accountability reform. The first changed the criterion from school performance to class performance, the second made the bonus criterion dependent purely on minority achievement, and the third made the bonus criterion tougher/easier to meet.

First, judging bonus receipt by individual performance increased the variance of bonus receipt, and divided teachers into two clear categories. While slightly more than half of the students benefit from the policy change, the average effort level of teachers and average scores statewide decline. Aggregate pay out per year would increase slightly, by about $1 million. However, the positive results must be weighed against the potential general equilibrium and administrative negative impacts and the fact that the policy increases the racial achievement gap.

Making the bonus depend on minority achievement had small but positive effects at the lower end of the achievement distribution (where most of the minority students are concentrated) and lowered teacher effort at higher achieving schools. Although the policy lowers average teacher effort and student performance statewide, there are gains made in minority-heavy schools. Through redistribution of effort, minority achievement in high performing schools may rise at the cost of white students. Aggregate pay out to teachers may decrease by about $4 million per year.

Finally, making the criterion tougher decreased teacher effort, led to overall decline in student achievement, and increased the racial achievement gap. Because the criterion was already too tough to be an efficient motivator of teachers, toughening the criterion made more teachers ‘give-up’ on the possibility of attaining the bonus. Making the standard easier to achieve actually increased effort, achievement, and reduced the achievement gap. However, there are limits to the policy’s effectiveness, as too lax a criteria then turns bonuses into a ‘sure-thing,’ inducing teachers to pull back on effort. Strengthening (weakening) the criterion by 10% would increase (decrease) aggregate bonus pay outs to teachers by about $5 million.
dollars per year.

While this paper attempted to bring more structure to measure teacher effort and evaluate the impact of accountability legislation, there are other factors that should be considered for a complete assessment of the incentive system. In particular, general equilibrium effects of teachers moving out of schools that fail repeatedly to attain the bonus should be considered. Other factors of interest include exploring how school administrators assign teachers and students to attempt to maximize the chances for bonus receipt.

One future research direction would be to make the model dynamic. The NCERDC data on teacher absences also records the month at which teachers took absences. Dividing the academic year into first and second semester, I can create a model in which the teacher receives a signal of her class/school’s predicted performance at the end of the first semester, allowing her to adjust her efforts accordingly in the second semester. See Clotfelter, Ladd, and Vigdor (2007) for more details.

The findings in this study hold generally for games in which managers must set all-or-nothing rewards for employees in a cooperative project with noisy output and signals. The current NC system is comparable to rewarding the team as a whole on the success of the overall project. The first simulation moves from team reward to individual contribution. The other two simulations change the standard by which success is judged: the second by isolating a particular aspect of the project as critical to success, and the third by making it more difficult/easier to achieve success. Ultimately, the research shows that, as is the case with most well-meaning legislation or incentive systems, there are always unintended responses from the targets of the legislation such that the end result may be quite different from what was originally intended.

References


Appendix

9.1 Proofs

Proof of existence of NE Since the utility function is concave in effort, the best response functions can be obtained by solving for first order conditions. Solving for first order conditions, \( J \) equations result:

\[
B \cdot F(\cdot) \frac{1}{J} \frac{\partial y_j}{\partial e_j} - C'(e_j) = 0
\]

By assumption, all teachers’ utility is strictly concave in her own effort. Re-write the first order conditions for teacher \( j \) as:

\[
B \cdot F(\cdot) \left( \frac{\alpha}{J} \left( e_j \cdot \exp(x_j) + \frac{1}{J-1} \sum_{k \neq j} e_k \cdot \exp(x_k) \right) \right) \frac{1}{J} \alpha \exp(x_j) = C'(e_j)
\]

Here, \( S_{-j} = \frac{\alpha}{J-1} \sum_{k \neq j} e_k \cdot \exp(x_k) \). Note that \( S_{-j} \in [e, \bar{e}] \) maps into \( e_j \in [e, \bar{e}] \) for all \( j \) (\( \bar{S} = \bar{e} \) by definition of \( S_{-j} \)).

Assume \( F''(\cdot) \geq 0 \), as we perturb \( S \) from its minimal to maximal value, \( e_j \) must increase. That is, the reaction function of \( e \) is positively sloped when \( F'' \geq 0 \). When the reaction function is positively sloped, \( e \) is a strategic complement, and it is well established that games in strategic complements have a unique pure strategy Nash equilibrium. See Milgrom and Shannon (1994) for details.

Assume \( F''(\cdot) \geq 0 \), as we perturb \( S \) from its minimal to maximal value, \( e_j \) must decrease. In this case, the reaction function is negatively sloped when \( F'' \leq 0 \). In this case, effort is a strategic substitute.

Multiple equilibria exist where 1) A subset of agents exerts effort, and 2) All agents exert some effort. If \( x^L < x_i < x^H \) and \( S_i < S_{-i} < S^H \) for \( i = 1, 2, ..., J \), all agents must exert some effort, and there exists a unique pure strategy Nash equilibrium. See Bramoulle and Kranton (2007) for details. QED.

Proof of Free Rider Problem The condition on \( x \) and \( S \) imply an interior solution. Since we assume that all teachers/classes are identical, \( \exp(x_j) e_j = \exp(x_{-j}) e_{-j} = S_{-j} \). Let \( x_{J+1} \), the additional class/teacher, also be identical to the other classes. Then, \( S_{-j} \) does not
change. The first order condition changes to:

\[
B \cdot F'(\frac{\alpha}{J+1} (e_j \cdot \exp(x_j) + \frac{1}{J} \sum_{-j} e_k \cdot \exp(x_k))) \cdot \frac{1}{J+1} \alpha \exp(x_j) = C'(e_j)
\]

which simplifies to:

\[
B \cdot F'(\alpha e \cdot \exp(x)) \cdot \frac{1}{J+1} \alpha \exp(x) = C'(e)
\]

This requires \(e\) to decrease in order for the FOC to be met. Since all teachers are identical, teachers reduce \(e\). The result follows that the incentive system suffers from a free rider problem. Note here that one cannot take derivatives to test for \(\frac{\partial e_j}{\partial J}\) because increases in \(J\) are discrete, and an increase in \(J\) also implies a new \(x_{J+1}\) and \(e_{J+1}\). QED.

**Proof of Individual Criterion Not Always Increasing Effort**

Let \(J = 2\). In individual criterion, the FOCs are:

\[
B \cdot F'(\alpha (e_1 \exp(x_1))) \alpha \exp(x_1) - C'(e_1) = 0 \\
B \cdot F'(\alpha (e_2 \exp(x_2))) \alpha \exp(x_2) - C'(e_2) = 0
\]

Let \(x_1 = x^H\) and \(x_2 = x^L\). By Assumption (3), \(\frac{\partial F'}{\partial e_i} \approx 0\) for \(i = 1, 2\). The solution is at a corner, and \(e_1^* = e_2^* = 0\).

In school-wide criterion, the FOCs are:

\[
B \cdot F'(\alpha (\frac{1}{2}(e_1 \exp(x_1) + e_2 \exp(x_2)))) \cdot \frac{1}{2} \alpha \exp(x_1) - C'(e_1) = 0 \\
B \cdot F'(\alpha (\frac{1}{2}(e_1 \exp(x_1) + e_2 \exp(x_2)))) \cdot \frac{1}{2} \alpha \exp(x_2) - C'(e_2) = 0
\]

\(\frac{\partial F'}{\partial e_i} > 0\). The FOCs are jointly satisfied if and only if \(e_1^* > 0\) and \(e_2^* > 0\). While \(e_1^* = e_2^* = 0\) is still a possible solution, the positive effort solution dominates, as \(U_i(e_i = 0) = 0\) for \(i = 1, 2\), but \(U_i(e_i \geq 0) \geq 0\) for \(i = 1, 2\). QED.

**Illustration of the Negative Relationship between Absence and Effort**

Assume teachers get a daily ‘potential effort’ draw from some distribution \(G(\delta) \in [0, \bar{\delta}]\), such that teacher \(i\) has \(\{\delta_{i1}, \delta_{i2}, ..., \delta_{id}, ..., \delta_{iD}\}\) where \(D\) is the total number of days in an academic year. Define \(\mu_i\) as the mean value of effort for teacher \(i\). There exists some level of energy \(\bar{\delta}\) such that if \(\delta_{id} < \bar{\delta}\), teacher \(i\) decides to take an absence for the day, \(A_{id} = 1\). Therefore, the probability of teacher \(i\) taking an absence on any given day is:

\[
Pr(A_{id} = 1) = Pr(\delta_{id} < \bar{\delta})
\]
Then, the number of absence days teacher $i$ takes in year $t$ is: $A_{it} = D \cdot Pr(\delta_{id} < \delta)$. Assuming that effort is not transferable day-to-day, potential effort on day $d$, $\delta_d$ is equivalent to actual effort $e_d$. For two teachers $i$ and $j$ where $\mu_i \geq \mu_j$, $A_{it} \leq A_{jt}$.

Now assume effort can be stored but decays from day-to-day at rate $\lambda$. For illustration purposes, I focus on the last two days of the academic year before the EOG exam, $D - 1$ and $D$. Education production during the two days is $\exp(x)e_{D-1} + \exp(x)e_D$ if the teacher teaches both days. If the teacher opts to take one day off and teach on the last day, education production is $\exp(x)0 + \exp(x)(\lambda e_{D-1} + e_D)$. If $\lambda < 1$, education production declines if effort is stored up. Therefore, if there is any effort decay, education production is maximized by teaching both days.

It is also possible that there is education decay from day-to-day as students forget material learned. Assume that education decays at rate $\xi$. Again, focusing on the last two days, education production if the teacher teaches both days is $\exp(x)\xi e_{D-1} + \exp(x)e_D$. If the teacher takes a day off, education production is $\exp(x)\xi 0 + \exp(x)(\lambda e_{D-1} + e_D)$. Teachers have no incentive to store up effort (and take extra absence) is $\lambda \geq \xi$.

To sum, if I assume there is no education or effort decay, the assumption $A_i \geq A_j$ if and only if $e_i \leq e_j$ always holds. If I assume effort can be stored, the condition holds if there is any decay in effort from day-to-day. If I assume that effort can be stored and education decays from day-to-day, the condition holds if the rate of decay of effort is greater than the rate of decay of education.
### 9.2 Tables

Table 1: Sample Statistics†

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (Std. Dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.507</td>
</tr>
<tr>
<td>Minority</td>
<td>0.352</td>
</tr>
<tr>
<td>Parent HS or less</td>
<td>0.729</td>
</tr>
<tr>
<td>Observations</td>
<td>931,419</td>
</tr>
<tr>
<td><strong>Teacher</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.077</td>
</tr>
<tr>
<td>Minority</td>
<td>0.151</td>
</tr>
<tr>
<td>Experienced</td>
<td>0.942</td>
</tr>
<tr>
<td>Certified</td>
<td>0.057</td>
</tr>
<tr>
<td>Absent days/Acad. Yr.</td>
<td>9.567 (6.110)</td>
</tr>
<tr>
<td>Observations</td>
<td>46,985</td>
</tr>
<tr>
<td><strong>Class</strong></td>
<td></td>
</tr>
<tr>
<td>% Male</td>
<td>0.508 (0.086)</td>
</tr>
<tr>
<td>% Minority</td>
<td>0.357 (0.271)</td>
</tr>
<tr>
<td>% Parent HS or less</td>
<td>0.696 (0.250)</td>
</tr>
<tr>
<td>Class size</td>
<td>22.70 (3.718)</td>
</tr>
<tr>
<td><strong>School</strong></td>
<td></td>
</tr>
<tr>
<td>School size</td>
<td>567.787 (196.987)</td>
</tr>
<tr>
<td>Number of teachers</td>
<td>37.063 (11.953)</td>
</tr>
<tr>
<td>Rural school</td>
<td>0.471</td>
</tr>
</tbody>
</table>

†NCERDC dataset. Years 2000 - 2004. All 3rd to 5th grade public elementary students and their teachers. Teachers with more than 30 days absence excluded. Students with one or no exam record excluded.
Table 2: First Stage Estimates†

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient (Std. Dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Number of Teachers</td>
<td>0.0531 (0.0234)</td>
</tr>
<tr>
<td>Log $\phi$</td>
<td>-0.1202 (0.0423)</td>
</tr>
<tr>
<td>Male</td>
<td>-0.149 (0.0143)</td>
</tr>
<tr>
<td>Minority</td>
<td>0.0482 (0.0108)</td>
</tr>
<tr>
<td>Experienced</td>
<td>0.8502 (0.0285)</td>
</tr>
<tr>
<td>Certified</td>
<td>0.1690 (0.0132)</td>
</tr>
<tr>
<td>Class size</td>
<td>-0.0020 (0.0010)</td>
</tr>
<tr>
<td>School size</td>
<td>-0.0002 (0.00004)</td>
</tr>
<tr>
<td>Rural school</td>
<td>0.0231 (0.0121)</td>
</tr>
<tr>
<td>% Class w/ parent HS or less</td>
<td>0.0206 (0.0146)</td>
</tr>
<tr>
<td>% Class minority</td>
<td>0.1014 (0.0191)</td>
</tr>
<tr>
<td>Observation</td>
<td>45,129</td>
</tr>
</tbody>
</table>

†Dependent variable is log of teacher absence in an academic year. Estimation also included district, year, and grade fixed effects. Errors are clustered at the school x year level.
Table 3: Structural and Comparison OLS Estimates of the Effect of Effort (Absence) on Student Achievement†

<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS Coefficient (Std. Dev.)</th>
<th>Structural Coefficient (Std. Dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Absence</td>
<td>-0.0071 (0.0011)</td>
<td>-0.2143 (0.0090)</td>
</tr>
<tr>
<td>Log Absence X Minority</td>
<td>0.0005 (0.0018)</td>
<td>0.0178 (0.0056)</td>
</tr>
<tr>
<td>Last year reading score</td>
<td>0.7096 (0.0007)</td>
<td>0.7094 (0.0007)</td>
</tr>
<tr>
<td>Male</td>
<td>-0.0367 (0.0013)</td>
<td>-0.0372 (0.0013)</td>
</tr>
<tr>
<td>Minority</td>
<td>-0.1602 (0.0042)</td>
<td>-0.1787 (0.0127)</td>
</tr>
<tr>
<td>Parent HS or less</td>
<td>-0.1919 (0.0018)</td>
<td>-0.1905 (0.0018)</td>
</tr>
<tr>
<td>Teacher certified</td>
<td>0.0282 (0.0028)</td>
<td>0.0631 (0.0031)</td>
</tr>
<tr>
<td>Teacher experienced</td>
<td>0.0854 (0.0040)</td>
<td>0.2189 (0.0066)</td>
</tr>
<tr>
<td>% Class w/ parent HS or less</td>
<td>-0.0405 (0.0038)</td>
<td>-0.0412 (0.0039)</td>
</tr>
<tr>
<td>% Class minority</td>
<td>-0.1064 (0.0040)</td>
<td>-0.0883 (0.0040)</td>
</tr>
<tr>
<td>Class size</td>
<td>-0.0010 (0.0002)</td>
<td>-0.0010 (0.0002)</td>
</tr>
<tr>
<td>School size</td>
<td>-1.86e-06 (4.12e-06)</td>
<td>-2.08e-05 (4.12e-06)</td>
</tr>
<tr>
<td>Observation</td>
<td>931,419</td>
<td>931,419</td>
</tr>
</tbody>
</table>

†Estimation also included district, year, and grade fixed effects. Student achievement is defined as standardized End-of-Grade reading exam scores.
Table 4: Individual Classroom and Race Specific Standards

<table>
<thead>
<tr>
<th></th>
<th>Status Quo</th>
<th>Classroom Performance</th>
<th>Minority Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected bonus</td>
<td>761.74 (81.48)</td>
<td>773.37 (263.61)</td>
<td>721.17 (127.86)</td>
</tr>
<tr>
<td>% Effort Increase</td>
<td></td>
<td>-4.18%</td>
<td>-3.56%</td>
</tr>
<tr>
<td>Δ Reading score†</td>
<td></td>
<td>-2.60</td>
<td>-2.06</td>
</tr>
<tr>
<td>% Teachers w/ higher effort</td>
<td></td>
<td>53.6%</td>
<td>10.3%</td>
</tr>
<tr>
<td>% Schools w/ higher scores</td>
<td></td>
<td>54.2%</td>
<td>10.9%</td>
</tr>
<tr>
<td>% Students w/ higher scores</td>
<td></td>
<td>52.8%</td>
<td>12.7%</td>
</tr>
<tr>
<td>Racial achievement gap‡</td>
<td>72.9</td>
<td>74.6</td>
<td>72.4</td>
</tr>
</tbody>
</table>

Among students/schools w/ higher scores

<table>
<thead>
<tr>
<th></th>
<th>Status Quo</th>
<th>Classroom Performance</th>
<th>Minority Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minority students</td>
<td>34.8%</td>
<td>33.4%</td>
<td>49.8%</td>
</tr>
<tr>
<td>Parents w/ HS or less</td>
<td>72.6%</td>
<td>70.6%</td>
<td>84.1%</td>
</tr>
<tr>
<td>School size</td>
<td>516.5</td>
<td>547.1</td>
<td>464.9</td>
</tr>
<tr>
<td>% Rural school</td>
<td>46.8%</td>
<td>44.6%</td>
<td>47.2%</td>
</tr>
</tbody>
</table>

†Change in percent of one standard deviation in test scores.
‡Percent of one standard deviation in test scores.
<table>
<thead>
<tr>
<th>Expected bonus</th>
<th>10% Tougher</th>
<th>10% Easier</th>
<th>20% Easier</th>
<th>30% Easier</th>
<th>40% Easier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status Quo</td>
<td>761.74 (81.48)</td>
<td>704.04 (82.70)</td>
<td>816.78 (80.04)</td>
<td>871.96 (78.46)</td>
<td>926.45 (76.75)</td>
</tr>
<tr>
<td>% Effort Increase</td>
<td>2.20%</td>
<td>2.20%</td>
<td>-1.30%</td>
<td>-1.30%</td>
<td>1.46%</td>
</tr>
<tr>
<td>∆ Reading score</td>
<td>-1.30</td>
<td>-1.30</td>
<td>-1.30</td>
<td>-1.30</td>
<td>-1.30</td>
</tr>
<tr>
<td>% Teachers w/ higher effort</td>
<td>87.6%</td>
<td>87.6%</td>
<td>87.6%</td>
<td>87.6%</td>
<td>87.6%</td>
</tr>
<tr>
<td>% Schools w/ higher scores</td>
<td>3.61%</td>
<td>3.61%</td>
<td>3.61%</td>
<td>3.61%</td>
<td>3.61%</td>
</tr>
<tr>
<td>% Students w/ higher scores</td>
<td>80.2%</td>
<td>80.2%</td>
<td>80.2%</td>
<td>80.2%</td>
<td>80.2%</td>
</tr>
<tr>
<td>Racial achievement gap†</td>
<td>72.9</td>
<td>72.9</td>
<td>72.9</td>
<td>72.9</td>
<td>72.9</td>
</tr>
<tr>
<td>% Teachers w/ higher effort</td>
<td>3.61%</td>
<td>3.61%</td>
<td>3.61%</td>
<td>3.61%</td>
<td>3.61%</td>
</tr>
<tr>
<td>% Schools w/ higher scores</td>
<td>80.2%</td>
<td>80.2%</td>
<td>80.2%</td>
<td>80.2%</td>
<td>80.2%</td>
</tr>
<tr>
<td>% Students w/ higher scores</td>
<td>72.9</td>
<td>72.9</td>
<td>72.9</td>
<td>72.9</td>
<td>72.9</td>
</tr>
<tr>
<td>Among students/schools w/ higher scores</td>
<td>72.9</td>
<td>72.9</td>
<td>72.9</td>
<td>72.9</td>
<td>72.9</td>
</tr>
<tr>
<td>Minority students</td>
<td>34.8%</td>
<td>34.8%</td>
<td>34.8%</td>
<td>34.8%</td>
<td>34.8%</td>
</tr>
<tr>
<td>Parents w/ HS or less</td>
<td>72.6%</td>
<td>72.6%</td>
<td>72.6%</td>
<td>72.6%</td>
<td>72.6%</td>
</tr>
<tr>
<td>School size</td>
<td>426.1</td>
<td>426.1</td>
<td>426.1</td>
<td>426.1</td>
<td>426.1</td>
</tr>
<tr>
<td>% Rural school</td>
<td>516.5</td>
<td>516.5</td>
<td>516.5</td>
<td>516.5</td>
<td>516.5</td>
</tr>
</tbody>
</table>

†Change in percent of one standard deviation in test scores.
‡Percent of one standard deviation in test scores.
Table 6: Alternative Specification of First Stage†

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient (Std. Dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of teachers</td>
<td>0.0091 (0.0043)</td>
</tr>
<tr>
<td>Expected bonus</td>
<td>-0.0227 (0.0047)</td>
</tr>
<tr>
<td>Expected bonus²</td>
<td>1.30e-05 (3.05e-06)</td>
</tr>
<tr>
<td>% Class minority</td>
<td>0.5747 (0.1507)</td>
</tr>
<tr>
<td>Male</td>
<td>-1.2804 (0.1045)</td>
</tr>
<tr>
<td>Minority</td>
<td>0.4299 (0.0834)</td>
</tr>
<tr>
<td>Experienced</td>
<td>4.4982 (0.1185)</td>
</tr>
<tr>
<td>Certified</td>
<td>1.6280 (0.1206)</td>
</tr>
<tr>
<td>Class size</td>
<td>-0.0091 (0.0072)</td>
</tr>
<tr>
<td>School size</td>
<td>-0.0011 (0.0003)</td>
</tr>
<tr>
<td>Rural school</td>
<td>0.3502 (0.0795)</td>
</tr>
</tbody>
</table>

†Dependent variable is teacher absence in an academic year. Estimation also included district, year, and grade fixed effects. Errors are clustered at the school x year level.
Table 7: Year-by-Year Regression of Free Rider Effect †

<table>
<thead>
<tr>
<th>Year</th>
<th>Coefficient (Std. Dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Incentive Policy</td>
<td></td>
</tr>
<tr>
<td>94-95</td>
<td>-0.0031 (0.0517)</td>
</tr>
<tr>
<td>After Incentive Policy</td>
<td></td>
</tr>
<tr>
<td>95-05</td>
<td>0.0473 (0.0116)</td>
</tr>
<tr>
<td>Year-by-Year</td>
<td></td>
</tr>
<tr>
<td>95-96</td>
<td>0.1640 (0.0771)</td>
</tr>
<tr>
<td>96-97</td>
<td>0.0589 (0.0242)</td>
</tr>
<tr>
<td>97-98</td>
<td>0.0180 (0.0525)</td>
</tr>
<tr>
<td>98-99</td>
<td>0.0149 (0.0373)</td>
</tr>
<tr>
<td>99-00</td>
<td>0.1127 (0.0304)</td>
</tr>
<tr>
<td>00-01</td>
<td>0.0323 (0.0321)</td>
</tr>
<tr>
<td>01-02</td>
<td>0.0511 (0.0252)</td>
</tr>
<tr>
<td>02-03</td>
<td>0.0190 (0.0393)</td>
</tr>
<tr>
<td>03-04</td>
<td>0.0518 (0.0341)</td>
</tr>
<tr>
<td>04-05</td>
<td>0.0624 (0.0311)</td>
</tr>
</tbody>
</table>

†Dependent variable is log of teacher absence in an academic year. The parameter estimate presented in the table is the log of number of teachers at the school. Estimation also included district, year (where appropriate), and grade fixed effects. Errors are clustered at the school x year level.
Table 8: Testing for Direct Effect of Teacher Exposure on Achievement†

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient (Std. Dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Absence</td>
<td>-0.2452 (0.0090)</td>
</tr>
<tr>
<td>Log Absence X Minority</td>
<td>0.0148 (0.0057)</td>
</tr>
<tr>
<td>Log Number of Teachers X Pr(Bonus) $\equiv 0$ or 1</td>
<td>-0.0029 (0.0035)</td>
</tr>
<tr>
<td>Last year reading score</td>
<td>0.7104 (0.0007)</td>
</tr>
<tr>
<td>Male</td>
<td>-0.0372 (0.0013)</td>
</tr>
<tr>
<td>Minority</td>
<td>-0.1866 (0.0130)</td>
</tr>
<tr>
<td>Parent HS or less</td>
<td>-0.1882 (0.0018)</td>
</tr>
<tr>
<td>Teacher certified</td>
<td>0.0700 (0.0031)</td>
</tr>
<tr>
<td>Teacher experienced</td>
<td>0.2434 (0.0066)</td>
</tr>
<tr>
<td>% Class w/ parent HS or less</td>
<td>-0.0415 (0.0040)</td>
</tr>
<tr>
<td>% Class minority</td>
<td>-0.0901 (0.0040)</td>
</tr>
<tr>
<td>Class size</td>
<td>-0.0010 (0.0002)</td>
</tr>
<tr>
<td>School size</td>
<td>-3.64e-05 (6.14e-06)</td>
</tr>
<tr>
<td>Observation</td>
<td>931419</td>
</tr>
</tbody>
</table>

†Estimation also included district, year, and grade fixed effects. Student achievement is defined as standardized End-of-Grade reading exam scores.
9.3 Figures

Figure 1: Non-cooperative Regime

Figure 2: Cooperative Regime
Figure 3: The effect of $\phi$ on effort exertion

![Graph showing the effect of $\phi$ on effort exertion.]

Figure 4: Illustrated Model

![Illustrated Model with points A, B, C, D, E, F, and G.]

 Effort

Student Achievement

Expected Bonus

Criterion Severity
Figure 5: Histogram of Expected Bonus under Status Quo and Individual Criterion

Figure 6: Histogram of Predicted Absence under Status Quo and Individual Criterion

Figure 7: Histogram of Predicted Test Scores under Status Quo and Individual Criterion
Figure 8: Best Response Functions for Status Quo and Minority-Only Standard