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

JOB MARKET PAPER

Tom Ahn†
Duke University

November 4, 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, NC public school data, and the state's unique accountability system that rewards teachers for 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 simulations to eliminate the free-rider problem and change the teacher's ability to affect her chance at bonus receipt, can yield higher teacher effort and student achievement. However, the appropriate level of incentives must be chosen with care, as poorly considered policy changes lead to lower achievement at higher cost compared to the status quo.

Keywords: Accountability, principal-agent model, teachers
JEL I21

1 Introduction

Nearly all education researchers and practitioners agree that teacher quality is one of the most important determinants of a student's success in his academic career. The literature has focused on identifying observable exogenous characteristics that proxy for teacher quality,

*I thank Peter Arcidiacono, Pat Bayer, Helen Ladd, Tom Nechyba, Stephen Ryan, Justin Trogdon, Jacob Vigdor, and seminar participants at Duke University. 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

finding that experience, education level, and credentials are correlated with higher student achievement.[1]

While these characteristics explain part of the effect of teachers, they fail to account for all of the observed variation in achievement. For econometricians, this residual unaccounted-for-effect of teachers has been labeled unobservable ability and modeled as individual-specific constants in fixed-effects models.[2] For policy makers, they continue to set up accountability legislation with sanctions or rewards aimed at teachers to raise student achievement.

Since accountability legislation is not designed to directly influence the observable characteristics mentioned above, it must be attempting to affect 'unobserved' ability. This paper uncovers this important and unexplored dimension of teacher quality as effort. Effort is the one endogenous characteristic common to all teachers that can be immediately influenced by incentives. Policy makers have used accountability systems such as the federal No Child Left Behind Act (NCLB) as blunt instruments to motivate teachers by introducing market forces.

This issue can be understood as a variant of the principal-agent problem.[3] The principal (government) is setting a wage contract (accountability incentives) 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). Free-rider issues exist because output is measured and rewarded at the school level.

I propose to exploit the behavioral response of teachers to incentives and distortions caused by free-rider effects to measure effort explicitly. This will allow me to analyze the intended and unintended effect of accountability legislation (efficiency as well as distributional implications), and the 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. Studies have shown that performance incentives for teachers can lead to higher academic performance.[4] States with stricter accountability standards have been associated with higher test scores (Carnoy and Loeb (2002)). High stakes testing improves student achievement, as compared to low stakes testing.[5] The accumulated evidence shows that incentives can change teacher behavior, helping to raise scores.

However, it is unclear whether the score increase reflects actual growth of education production. Student achievement can be raised in one of two ways. The first way is to use underhanded means to artificially inflate scores.[6] This introduces noise into the system, mak-

[1]See Goldhaber and Anthony (2007), Rockoff (2004), Clotfelter, Ladd and Vigdor (2007), among others.

[2]See Rockoff (2004), Rivkin, Haushek, and Kain (2005).

[3]See Sappington (1991) for a review.

[4]See Figlio and Kenny (2006), Clotfelter and Ladd (1996), Ladd (1999), Eberts, Hollenbeck, and Stone (2002).

[5]See Vigdor (2008) and Jacob (2007).

[6]Examples of these behaviors include classifying marginal students as disabled or suspending them to prevent

ing exam scores a poor measure of student achievement. The second way is for teachers to exert more effort. An increase in effort will lead to real education production (assuming effort matters).

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 little evidence on how many dollars are required to raise teacher effort by enough to affect student performance by a targeted amount. Furthermore, it is clear that incentives will not affect all teachers equally. Teachers closer to whatever threshold the government sets may behave differently compared to teachers farther away. This study will provide a means to estimate the differential causal effects. Differences in incentive strength and free-rider effects induce measurable behavioral changes in teachers that are unrelated to underhanded actions.

Furthermore, recent research has focused on the distributional changes in achievement.[7] Besides the change in distribution of resources/effort, which clearly defines winners and losers among the student 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 students in the public school system. This dataset, along with the unique teacher incentive system that pays out bonuses for school-level year-over-year improvement in 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 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. As the number of teachers increases (while the size of individual rewards stays constant), each teacher's optimal effort declines. 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 chance 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

---

testing (Cullen and Reback (2002), Figlio and Getzler (2002), Jacob (2005), Figlio (2006)), 'teaching to the test' (Grissmer and Flanagan (1998), Hanushek and Raymond (2004)), and cheating by *teachers* who alter students' answer sheets (Jacob and Levitt (2003)).

[7] See Neal and Schanzenbach (2007), Booher-Jennings (2005), and White and Rosenbaum (2007).

achievement, white students are more sensitive to effort compared to minority students, and teachers adjust effort to maximize utility. Using parameter estimates, I run policy simulations that 1) change the bonus regime from cooperative (school average criterion) to non-cooperative (classroom or grade-level average criterion), and 2) make the criterion tougher or easier to reach. The first policy change tries to reduce the free-rider problem intrinsic to the incentive policy; the second change aims to induce higher teacher effort by changing the teacher's ability to affect her own probability of bonus receipt.

The simulation results show that setting arbitrary performance standards and hoping for the best is naive. Attempting to completely eliminate free-rider effects by evaluating bonus at the classroom level costs more and leads to average achievement actually declining. A better solution is to evaluate at the grade-level, which leads to achievement increase at zero cost. However, changing the system may impose other general equilibrium and administrative problems. Attempting to induce higher effort by making the bonus tougher to get leads to lower achievement. Laxer standards in fact lead to higher scores, but only up to a point. Making the bonus too easy to attain results in lower scores. Unfortunately, easier standards make the incentive system substantially more expensive.

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 econometric 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 academic year. While the system has grown in complexity and has gone through minor alterations in the details of execution,[8] 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' growth are defined below.

[8]For 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).

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

$$\Delta y_{mgst} = \Delta y_{gs94} + b_1 ITP_{mgt} + b_2 IRM_{mgst}$$

The Δy_{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. Δy_{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 ITP_{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 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[9], 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 fail 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.

Figure 1 shows the change in average End-of-Grade exam scores from 1993 to 2006. The vertical line represents the year at which the incentive policy was introduced. To note is that the thick line and the dashed line, which represent pre-incentive scores and post-incentive scores respectively, grow at different rates. After the ABC policy is introduced, scores grow at a faster rate, implying that teacher behavior (or the test itself) has changed.

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. Teachers are differentiated by two dimensions, ability, $x_j \in [\underline{x}, \overline{x}]$, and effort, $e_j \in [\underline{e}, \overline{e}]$.[10] B is the bonus that is paid to all teachers at the school upon qualifying under the state criterion. Cr

[9]The difference is divided by the standard deviation of the difference across all schools in the state.

[10]I abstract away from student ability in the theory section, but the econometric model controls for observable student and peer demographic characteristics.

is the state defined criterion that the school must beat in order to qualify for the bonus.

$$U_j = B \cdot Pr(e_1, e_2, ..., e_J, x_1, x_2, ..., x_J, Cr) - C(e_j) \tag{1}$$

The point to emphasize here is that bonus receipt is determined at the school level. Therefore, all teachers' efforts and abilities 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 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:

$$y_{jt} = a(x_{jt}, e_{jt}) \tag{2}$$

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 probability of success is defined by:

$$Pr_t = F\left(\frac{\sum_1^J a(x_{kt}, e_{kt})}{J} - \overline{y}_{t-1} > Cr\right) \tag{3}$$

I make the following functional form assumptions:

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

$$a(x_{jt}, e_{tj}) = \exp(x_{jt})e_{jt}$$

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

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

4. Define $S_{-j} \in [\underline{S}, \overline{S}]$ 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, S^L, such that:

$$F(e_j|S_{-j} \geq S^H) \rightarrow 1 \ , \ F(e_j|S_{-j} \leq S^L) \rightarrow 0 \qquad \forall e_j,\, x_j$$

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

$$\frac{\partial F(\cdot|x_j \geq x^H)}{\partial e} \cong 0 \ , \ \frac{\partial F(\cdot|x_j \leq x^L)}{\partial e} \cong 0$$

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 4 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 (and/or are inept), the best response of teacher j is also to shirk. On the other hand, if all teachers are giving maximal effor (and/or are highly-capable) 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 5 is a statement about the limitations of effort within the classroom. That is, if the teacher's ability is very low (or very high), she cannot substantially increase her probability of receiving the bonus. When teacher ability is very high, the probability of students meeting the state criterion approaches one, and additional effort from the teacher has little impact. Similarly, additional effort fails to make up for a teacher with very low ability, and the probability of receiving the bonus remains 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,* $x^L < x < x^H$*, 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 $J + 1$.[11] As J increases, a teacher's effort is distributed over a larger population. Since her share remains constant, she will be induced to lower her level of effort. In this sense, free-rider problems always exists.

[11] Note that in this context, an increase in J is a discrete increase. $J + \epsilon$ is a meaningless concept here.

There is a second, indirect channel to consider. An increase in the number of teachers in the model necessarily implies an additional class.[12] This means that x_{J+1} is included in the probability of receiving the bonus, which can change the distribution of ability of teachers in the school. In proposition 2, I assumed that the distribution of ability does not change to focus on the direct channel. However, it is easy to imagine a scenario where a new teacher's 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:

$$J^{-1} \cdot B \cdot F'(y_m) \cdot \exp(x_j) = C'(e_j)$$

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 γ 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 (cooperative) incentive to a classroom-level (non-cooperative) incentive where teachers are judged purely on her students' performance. The first order condition shows that the $-\ln J$ term 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.*

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 her 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 changing the probability of bonus receipt. This change in optimal effort exertion based on a teacher's placement on the pdf of the bonus receipt is critical in the analysis of any policy that attempts to increase student achievement by altering the measurement method of the incentive system.[13]

[12] I assume away the possibility of J increasing and the average size of classes decreasing.
[13] I call this the 'location' effect as a short-hand for exposition in later sections.

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 2 and Figure 3. 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 4), their optimum solution is at the corner: $e_1 = e_2 = 0$. Now, assume that bonus is awarded for joint performance.[14] 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.

The theory results indicate that a collaborative regime that experiences some free-rider problems may be preferable to a non-cooperative regime depending on the distribution of ability. If schools are composed mostly of homogeneous teachers, 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 composed of roughly the same number of very high and very low ability teachers, a cooperative incentive policy may actually induce greater effort and increased achievement.[15]

The school-wide criterion can be considered as a blunt instrument to gather teachers into a narrower band on the bonus receipt probability. If teachers are sparsely and evenly spaced out across the probability distribution, the incentive system will not matter to the majority of teachers. When success depends upon joint performance, teachers at the left and right extremes are 'gathered' to the middle of the distribution. The tighter the band on the distribution is, the more teachers are potentially affected by the incentive policy. This potential positive effect of joint performance criterion must be weighed against the free-rider effects inherent to the problem.

Tightly gathering teachers will only increase the power of the incentives if the spot at which teachers are gathered allows them to positively affect their probabilities through the application of effort. One can imagine changing Cr to get teachers to the middle of the pdf, where the marginal effects are the largest. The policy simulation section explores these theoretical implications more fully.

[14] I 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.

[15] Note that cooperation here does *not* imply the usual context of sharing knowledge or explicitly working together in any fashion. It only refers to the fact that success is judged on joint achievement.

4 Data

I use an administrative dataset for the North Carolina public school system from the academic years 1999-00 to 2003-04. [16] 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 should 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) 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.[17]

Because sick and personal leaves are unplanned-for and take place during the school year, they have the most effect on student learning. Moreover, teachers willing to take days off

[16] The 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). Post-2004 data has unreliable absence data, and pre-1999 data yield poorer teacher-student matches in the dataset.

[17] About 60% of annual leaves are concentrated in December, June, July, and August. Another 20% is taken in November, which indicates Thanksgiving break.

during the school year know it can negatively impact her students, as a substitute will have to fill in, potentially leading to lower quality instruction for that day. For the study period, teachers took an average of 9.6 days of absence 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 absence in the academic year. There are significant numbers of teacher who take more than 30 days off (about 5% of the original sample). While not explicitly measured in the dataset, I believe that a significant portion of these long-term absences are maternity leave. Over 90% of the elementary school teacher population is female. I exclude these teachers from analysis.[18] The resulting sample absence rate of 5.3% is in line with other studies of teacher absences. [19] Table 1 summarizes student, teacher, class, and school characteristics and absence taking behavior.

4.1 Reduced-form Results

Before moving to the econometric model, I demonstrate here that 1) the incentive policy did lead to behavioral changes by teachers, and 2) free-rider effects are readily observable in the data. Table 2 shows the average number of absences for elementary school teachers from 1995-1996 to 2003-2004. The incentive system was put in place in the 1996-1997 school year, so the first year can be considered the base-line before teachers were offered cash bonuses. A decline in the average absence rate is clearly observable after the introduction of the incentive policy. While other macro shock as well as the strength of the incentive criteria drive the absence rate up and down, it is always lower compared to the pre-incentive regime.

Table 3 shows results from regressing test scores on the number of teachers for academic years 1999-2000 to 2003-2004. The negative and significant parameter estimate on the number of teachers points to the existence of free-rider effects in the North Carolina incentive system. In fact, this result is even stronger, as it is possible to argue that a larger number of teachers should have indirect positive effects.[20]

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,

[18] Including teachers with high level of absence does not qualitatively change results.

[19] See Ehrenberg et. al. (1991).

[20] For instance, the student may have more academic resources, discipline may be easier to uphold with more teachers, etc. This possibility is further explored in the results section.

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.[21]

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 $\widehat{y_{st}}$ represent the predicted *school* performance this year, and let $y_{s(t-1)}$ be school performance last year. Expected bonus is simple to calculate:

$$E(B) = \{\Phi(\widehat{y_{st}} - y_{s(t-1)} \geq Cr_1) + \Phi(\widehat{y_{st}} - y_{s(t-1)} \geq Cr_2)\} \cdot B \tag{4}$$

where Cr_1 and Cr_2 are expected and high growth composite threshold values, respectively, and B is the bonus amount.[22] 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:

$$e_{jst} = \alpha_0 + X_{jst}\alpha_1 + \alpha_2 I_{jst} + \varepsilon_{jst}$$

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 ε_{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 contains 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:

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

[21] This is equivalent to solving for a teacher reaction function. I attempted an alternative specification of estimating student achievement non-parametrically based purely on student demographic information to plug into the school-level expected bonus equation. This negates the need for iteration to solve for the reaction function. This made no qualitative difference in the results.

[22] Note that in the estimation, following the first order condition, I use the pdf value $\phi(\widehat{y_{st}} - y_{s(t-1)} \geq Cr_1) + \phi(\widehat{y_{st}} - y_{s(t-1)} \geq Cr_2)$.

η_{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 orthogonal shocks that affect the teacher's absence outcome.

I assume that for two teachers i and j, $A_i \geq A_j$ if and only if $e_i \leq e_j$. 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 in the academic 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, and saturate her students at exam time.[23]

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.[24] Using the data available and the first order condition directly from the theory, I model the absence decision as follows:[25]

$$\ln A_{jst} = X_{jst}\alpha_1 + \alpha_2 \ln J_{st} + \alpha_3 \ln(\phi(\hat{y}_{st}, Cr_1) + \phi(\hat{y}_{st}, Cr_2)) + \eta_{jst} \tag{5}$$

The measure of incentive strength (I_{jst}) to be used are J_{st}, the number of teachers at the school, and $\phi(\hat{y}_{st}, Cr_1) + \phi(\hat{y}_{st}, Cr_2)$, 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 ϕ term is best explained by Figure 4. The ϕ 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 mid-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 ϕ variable is important, as it reflects the non-linear effect of the probability of receiving the bonus. As bonus becomes very easy or very difficult to attain, the utility maximizing response of teachers is to decrease their effort. The non-linear, non-monotonic effect will be important for policy analysis.[26] η_{jst} is the idiosyncratic error term that represents the uncorrelated health and weather shocks.

[23] See the appendix for a formalization of this condition.

[24] This relationship relies on the assumption that η is uncorrelated with I. That is, higher or lower incentive strength is uncorrelated with good or bad health/weather shocks.

[25] $A_{jst} = 0$ composes less than 1.5 % of the sample. Setting those values to 0.01 and inserting them into the sample made no qualitative difference in the results.

[26] An alternative specification may shed light on the role of ϕ. To check to see if the shape of the pdf is consistent with the theory, I replace ϕ with a monotone measure $E(B)$, the expected bonus.:

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

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 4, and results support the functional form assumption.

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 *per se* is not rewarded as part of the incentive system.[27] $\widehat{\ln A}$ is now a latent variable for effort.

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. The achievement function for student i is specified as follows:

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

Z_{ist} is a vector of student, peer, and teacher demographic characteristics. $\widehat{A_{jst}}$ is the projected measure of (negative) effort from equation (1). ν_{ist} is the idiosyncratic error. The reason for the inclusion of $\widehat{A_{jst}}$ is to estimate the effect of year-long teacher effort on student achievement.[28]

I make a concession to expediency[29] in estimating the achievement function by including lagged achievement. Lagged achievement does not appear in the theoretical model, and the inclusion of $y_{is(t-1)}$ creates econometric 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.[30] The functional form in this study assumes a constant rate of decay across all students and grades. An alternate form of

[27]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.

[28]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 *positive* effect on achievement, I run an alternative specification using $J \cdot I[\text{no bonus effect}]$ as an exclusion 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 5.

[29]Another concession is the inclusion of the interaction term of log of absence and an indicator for minority student. While the achievement function defined in the theoretical section implies effort is completely separable, the difference in effort effectiveness across races is important for teacher utility optimization as well as the evaluation of effectiveness of suggested policy changes.

[30]Differencing y_{ist} and $y_{is(t-1)}$ in the dependent variable and estimating a 'gains' equation yielded no qualitative differences. However, this assumes that there is zero educational decay from one year to the next (after students take summer break).

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.[31]

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.[32]

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 7. 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.[33] 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 estimate of the impact of projected effort due to incentive effects is about 30 times greater than the OLS parameter estimate on absence.[34] This large difference can be explained by extending the example above. Assume that 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 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

[31] See Rothstein (2008).

[32] Alternatively, it is possible that effort is not important. Pure exposure to teachers may drive achievement.

[33] 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.

[34] Including fixed-effect parameters at the student level yielded no qualitative differences.

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 average effect across the two teachers). 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 projected effort are fundamentally different measures. Each incremental increase in projected effort is reflected in the quality of instruction for the *entire* academic year.

The first-stage estimates in Table 6 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 in school-level incentives[35].

The sign of the parameter on the ϕ term shows that the *location* of a teacher on the pdf of the bonus receipt is critical in determining her effort exertion. Very low and very high probabilities of bonus receipt are associated with low effort exertion. Translating the ϕ 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 \$870, 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[36].

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.

[35] 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 8.

[36] The estimation is not quasi-experimental in that I do not have an explicit measure for pre-incentive and post-incentive effects. This is due to pre-incentive data limitations, as only about 25 to 30% of the student data can be matched across two pre-incentive years. Given this limitation, I run a placebo trial that estimates the teacher absence decision equation and the effect of incentives on achievement. The results show that teachers do not respond and student achievement is not driven by incentives (as they do not exist yet). See Table 9 and Table 10.

It is interesting to place these results in the context of the controversy over the effectiveness of class size reduction (See Hanushek, Rivkin, and Kain (2005)). While magnitudes differ by studies, it is generally agreed that reducing the class size by approximately 10 students may yield somewhere between 5 to 15% of a standard deviation increase in student achievement. The benefits of increasing expected bonus amounts falls somewhere in between the low-end and high-end of these estimates. The cost of splitting a classroom, which will require an additional full-time teacher plus classroom space, will be much higher than doubling the teacher's expected bonus amount.

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 5. 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 two possible policy changes. The first experiment changes the criterion from a school-wide performance to classroom or grade-level performance measure, attempting to eliminate the incentive dilution problem. The second 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 counterproductive. I present two possible policy alterations to the current system to examine their possible effects on teachers and students. The first policy change attempts to raise achievement by reducing the direct impact of free-ridership. This is done by targeting bonus at the classroom or grade level. The second policy attempts to fine-tune the location of teachers on the probability of distribution to induce higher effort. This is done by making the bonus tougher or easier to attain.

7.1 Targeting at Classroom or Grade Level

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, one may predict lower effort as the number of teachers in the school (controlling for student population) increases.

Further, 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, fix the number of teachers in the school to one, and predict effort. The results are shown in the second column of Table 11.

The results from the simulations show that expected bonus on average increases slightly.[37] 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. As a result of the decrease in effort, state average reading score declines.

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 ability and high ability teachers.[38] 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

[37]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.

[38]In the data, 'low' ability 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.

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.

Therefore, moving from school-wide to classroom standards completely eliminates the direct effect of free-ridership, but increases the indirect effect arising from locating more teachers at the tails of the pdf. The histograms in Figures 6, 7, and 8 demonstrate this indirect effect. Figure 6 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 7 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 of teachers exerting low effort increases. This is directly attributable to the increase in the tails of the first histogram. Highly under/over-achieving teachers are no longer pushed toward the middle of the distribution, 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 8.

If attempting to completely eliminate the direct free-rider effect through individual incentives results in negative achievement growth due to the location effect, an intermediate step of targeting by grade-level may mitigate the direct effect (through reduction in number of teachers), yet retain the desirable cooperative properties of the indirect, location effect. The third column in Table 11 demonstrates this to be the case.

Expected bonus pay out per teacher is virtually unchanged from the status quo, yet average teacher effort and student achievement increase. The differences in variance across the three regimes show that targeting at the grade-level substantially lowers the fraction of teachers pushed out to the tails of the expected bonus distribution compared to classroom-level targeting. At the same time, the reduction in the number of teachers by approximately a third, compared to the school-level targeting, raises the base effort level of all teachers. The combination of these two effects manages to hit a 'sweet-spot,' where it is possible to have a virtually costless increase in teacher effort exertion and student achievement growth. Comparing the racial achievement differences across the three regimes show that classroom-level targeting is regressive and grade-level targeting is mildly progressive. Therefore, the beneficiaries of individual incentives would be white students, and minority students would benefit from grade-level incentives.

The policy simulation results would seem to indicate that the state should immediately switch to a grade-level criterion. However, several problems arise with switching to the proposed standard. There could be large, negative general equilibrium effects. For instance, teachers may compete to avoid grades 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. In addition, 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 an alternative criterion, these negative factors must be considered carefully, along with the positive results from the mitigation of free-ridership.

7.2 Tougher or Easier Bonus Requirements

Targeting at the classroom or grade-level attempts to increase achievement by decreasing the direct effect of free-ridership. An alternative method would be to control the location effect, by changing the severity of the bonus criterion. In essence, by altering how easy or difficult it would be to get the bonus, a teacher's position on the probability distribution would change.

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. 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 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 12.

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 12 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.[39]

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 achievement gap was examined. Distilling the absence taking behavior of teachers into a signal for effort and using 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 compared to their minority counterparts, and teachers respond to this difference in sensitivity by altering their effort exertion.

I performed two policy simulations to gauge the possible effects of accountability reform.

[39]Unlike 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.

The first experiment changed the criterion from school to class or grade-level performance, attempting to raise achievement by eliminating or mitigating the direct free-rider effect. The second experiment made the bonus criterion tougher or easier to meet, attempting to raise achievement by fine-tuning the location effect.

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. The intermediate policy of judging by grade increased overall achievement and decreased the racial achievement gap at zero cost. However, the positive results must be weighed against the potential general equilibrium and administrative negative impacts.

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 simulation changes the standard by which success is judged, 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

[1] Arcidiacono, P., G. Foster, N. Goodpaster, J. Kinsler (2007) "Estimating Spillovers using Panel Data, with an Application to the Classroom." Duke University, working paper.

[2] Booher-Jennings, J. (2005) "Below the Bubble: 'Education Triage' and the Texas Accountability System." American Educational Research Journal. (42): 231-268.

[3] Bramoulle Y., R. Kranton (2007) "Public Goods in Networks." Journal of Economic Theory (135) : 478-494.

[4] Carnoy M., S. Loeb (2002) "Does External Accountability Affect Student Outcomes? A Cross-State Analysis" Educational Evaluation and Policy Analysis, 24: 305-331.

[5] Charkabarti, R. (2006) "Vouchers, Public School Response and the Role of Incentives: Evidence from Florida." Federal Reserve Bank of New York, mimeo.

[6] Clotfelter C., H. Ladd (1996) "Recognizing and Rewarding Success in Public Schools." In: Ladd, H. (Ed.), Holding Schools Accountable: Performance-Based Reform in Education. Brookings Institution, Washington, D.C..

[7] Clotfelter, C., H. Ladd, J.Vigdor and R. Aliaga. (2004) "Do School Accountability Systems Make It More Difficult for Low Performing Schools to Attract and Retain High Quality Teachers?" Journal of Policy Analysis and Management 23: 251-271.

[8] Clotfelter, C., H. Ladd, J. Vigdor (2006) "Teacher-Student Matching and the Assessment of Teacher Effectiveness." NBER Working Paper 11936.

[9] Clotfelter, C., H. Ladd, J. Vigdor (2007) "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects." NBER Working Paper 13617.

[10] Clotfelter, C., H. Ladd, J. Vigdor (2007) "Are Teacher Absences Worth Worrying about in the U.S.?" NBER Working Paper 13648.

[11] Clotfelter, C., H. Ladd, J. Vigdor (2008) "The Academic Achievement Gap in Grades 3 Through 8." Forthcoming in Review of Economics and Statistics.

[12] Cook, P, R. MacCoun, C. Muschkin, J. Vigdor (2008) "The Negative Impacts of Starting Middle School in Sixth Grade." Journal of Policy Analysis and Management 27 v.1.

[13] Cooley, J. (2008) "Desegregation and the Achievement Gap:Do Diverse Peers Help?" Working Paper.

[14] Cullen, J., R. Reback (2002) "Tinkering Toward Accolade: School Gaming under a Performance Accountability System." University of Michigan Working Paper.

[15] Dubey, P. O. Haimanko, A. Zapechelnyuk (2006) "Strategic Complements and Substitutes, and Potential Games." Games and Economic Behavior 54: 77-94.

[16] Ehrenberg, R., R. Ehrenberg, D. Rees and E. Ehrenberg (1991) "School District Leave Policies, Teacher Absenteeism, and Student Achievement," Journal of Human Resources (26): 72-105.

[17] Figlio, D., L. Getzler (2002) "Accountability, Ability, and Disability: Gaming the System?" NBER Working Paper 9307.

[18] Figlio, D. (2006) "Testing, Crime, and Punishment." Journal of Public Economics 90: 837-851.

[19] Figlio, D., C. Rouse (2006) "Do Accountability and Voucher Threats Improve Low-Performing Schools?" Journal of Public Economics 90: 239-255.

[20] Grissmer, D., A. Flanagan (1998) "Exploring Rapid Achievement Gains in North Carolina and Texas." National Education Goals Panel.

[21] Hanushek E., M. Raymond (2002) "Lessons about the Design of State Accountability Systems." Paper prepared for *Taking Account of Accountability: Assessing Policy and Politics.*

[22] Hanushek E., M. Raymond (2005) "Does School Accountability Lead to Student Improvements?" Journal of Policy Analysis and Managment, 24:297-327.

[23] Hoxby C., G. Weingarth (2005) "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects." Working Paper.

[24] Jacob B., (2005) "Accountability, Incentives and Behavior: Evidence from School Reform in Chicago." Journal of Public Economics 89: 761-796.

[25] Jacob B., (2007) "Test-Based Accountability and Student Achievement: An Investigation of Differential Performance on NAEP and State Assessments." NBER Working Paper 12817.

[26] Kinsler J. (2007) "Suspending the Right to an Education or Preserving It? A Dynamic Equilibrium Model of Student Behavior, Achievement, and Suspension." University of Rochester, Working Paper.

[27] Lazear E. (2001) “Educational Production.” Quarterly Journal of Economics (116) :777-803.

[28] Milgrom P., C. Shannon (1994) “Monotone Comparative Statics.” Econometrica (62) : 157-180.

[29] Nechyba T. (2003) “Centralism, Fiscal Federalism, and Private School Attendance.” International Economic Review (44) : 179-204.

[30] Neal D., D. Schanzenbach (2007) “Left Behind by Design: Proficiency Counts and Test-Based Accountability.” NBER Working Paper 13293.

[31] Rivkin S., Hanushek, E., and J. Kain (2005) “Teachers, Schools, and Academic Achievement.” Econometrica (73) : 417-458.

[32] Rothstein J. (2008) “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement.” Working Paper.

[33] Rouse C., J. Hannaway, D. Goldhaber, and D. Figlio (2007) “Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure.” NBER Working Paper W13681.

[34] Sappington D. (1991) “Incentives in Principal-Agent Relationships.” The Journal of Economic Perspectives (5) : 45-66.

[35] Vigdor, J. (2008) “Teacher Salary Bonuses in North Carolina.” National Center for Performance Incentives Working Paper 2008-03.

[36] West, M. and P. Peterson (2006) “The Efficacy of Choice Threats within School Accountability Systems: Results from Legislatively Induced Experiments.” The Economic Journal, 116:C46-C62.

[37] White K., J. Rosenbaum (2007) “Inside the Blackbox of Accountability: How High-Stakes Accountability Alters School Culture and the Classification and Treatment of Students and Teachers.” Forthcoming in No Child Left Behind and the Reduction of the Achievement Gap: Sociological Perspectives on Federal Education Policy. Eds A. Sadvonik, J. O’Day, G. Bornstedt, and K. Borman. Routledge.

9 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:

$$\begin{aligned} B \cdot F'(\cdot)\frac{1}{J}\frac{\partial y_j}{\partial e_j} - C'(e_j) &= \\ B \cdot F'(\cdot)\frac{1}{J}\alpha\exp(x_j) - C'(e_j) &= 0 \end{aligned}$$

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

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

Here, $S_{-j} = \frac{\alpha}{J-1}\sum_{-j} e_k \cdot \exp(x_k)$. Note that $S_{-j} \in [\underline{e}, \overline{e}]$ maps into $e_j \in [\underline{e}, \overline{e}]$ for all j ($\overline{S} = \overline{e}$ 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 $S^L < S_{-i} < S^H$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:

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

which simplifies to:

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

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:

$$\begin{aligned} B \cdot F'(\alpha(e_1 \exp(x_1)))\alpha \exp(x_1) - C'(e_1) &= 0 \\ B \cdot F'(\alpha(e_2 \exp(x_2)))\alpha \exp(x_2) - C'(e_2) &= 0 \end{aligned}$$

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

In school-wide criterion, the FOCs are:

$$\begin{aligned} B \cdot F'(\alpha(\frac{1}{2}(e_1 \exp(x_1) + e_2 \exp(x_2))))\frac{1}{2}\alpha \exp(x_1) - C'(e_1) &= 0 \\ B \cdot F'(\alpha(\frac{1}{2}(e_1 \exp(x_1) + e_2 \exp(x_2))))\frac{1}{2}\alpha \exp(x_2) - C'(e_2) &= 0 \end{aligned}$$

$\frac{\partial F'}{\partial e_i} > 0$. 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 $G(\delta) \in [0, \overline{\delta}]$, 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 μ_i as the mean value of effort for teacher i. There exists some level of energy $\underline{\delta}$ such that if $\delta_{id} < \underline{\delta}$, 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:

$$Pr(A_{id} = 1) = Pr(\delta_{id} < \underline{\delta})$$

Then, the number of absence days teacher i takes in year t is: $A_{it} = D \cdot Pr(\delta_{id} < \underline{\delta})$. Assuming that effort is not transferable day-to-day, potential effort on day d, δ_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 λ. 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 decays from day-to-day as students forget material learned. Assume that education decays at rate ξ. 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 and Figures

Table 1: Sample Statistics†

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

†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: Average Days of Absence of Elementary School Teachers by Year

Year	Absence
1995-1996	9.9234
1996-1997	9.4088
1997-1998	9.4537
1998-1999	9.4987
1999-2000	9.8543
2000-2001	9.6498
2001-2002	9.9094
2002-2003	8.9368
2003-2004	8.5703

Table 3: Testing for Free-Rider Effects using Reduced Form†

Variable	Coefficient (Std. Dev.)
Log Number of Teachers	-0.1084 (0.0357)
Last year reading score	0.7363 (0.0008)
Male	-0.2866 (0.0129)
Minority	-1.7060 (0.0169)
Gifted Class	0.8000 (0.1333)
% Class minority	-1.3992 (0.0395)
Class Size	-0.0009 (0.0021)
Teacher male	-0.2008 (0.0248)
Teacher minority	-0.0331 (0.0202)
Teacher certified	0.284 (0.0298)
Teacher experienced	0.6749 (0.0259)
School size	-4.00e-04 (6.06e-05)
Rural school	-0.2049 (0.0183)
Observation	931,419

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

Table 4: Alternative Specification of First Stage†

Variable	Coefficient (Std. Dev.)
Number of teachers	0.0091 (0.0043)
Expected bonus	-0.0227 (0.0047)
Expected bonus^2	1.30e-05 (3.05e-06)
% Class minority	0.5747 (0.1507)
Male	-1.2804 (0.1045)
Minority	0.4299 (0.0834)
Experienced	4.4982 (0.1185)
Certified	1.6280 (0.1206)
Class size	-0.0091 (0.0072)
School size	-0.0011 (0.0003)
Rural school	0.3502 (0.0795)

†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 5: Testing for Direct Effect of Teacher Exposure on Achievement†

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

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

Table 6: First Stage Estimates†

Variable	Coefficient (Std. Dev.)
Log Number of Teachers	0.0531 (0.0234)
Log ϕ	-0.1202 (0.0423)
Male	-0.149 (0.0143)
Minority	0.0482 (0.0108)
Experienced	0.8502 (0.0285)
Certified	.1690 (0.0132)
Class size	-0.0020 (0.0010)
School size	-0.0002 (0.00004)
Rural school	0.0231 (0.0121)
% Class w/ parent HS or less	0.0206 (0.0146)
% Class minority	0.1014 (0.0191)
Observation	45,129

†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 7: Structural and Comparison OLS Estimates of the Effect of Effort (Absence) on Student Achievement†

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

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

Table 8: Year-by-Year Regression of Free Rider Effect †

Year	Coefficient on J (Std. Dev.)
Before Incentive Policy	
95-96	-0.0031 (0.0517)
After Incentive Policy	
96-05	0.0473 (0.0116)
Year-by-Year	
96-97	0.1640 (0.0771)
97-98	0.0589 (0.0242)
98-99	0.0180 (0.0525)
99-00	0.0149 (0.0373)
00-01	0.1127 (0.0304)
01-02	0.0323 (0.0321)
02-03	0.0511 (0.0252)
03-04	0.0190 (0.0393)
04-05	0.0518 (0.0341)
05-06	0.0624 (0.0311)

†Dependent variable is log of teacher absence in an academic year. The parameter estimate presented in the table is the log of number of teacers at the school Estimation also included teacher and class demographic variables and district, year (where appropriate), and grade fixed effects. Errors are clustered at the school x year level.

Table 9: First Stage Estimates Using Pre-Incentive Data

Variable	Coefficient (Std. Dev.)
Number of Teachers	-0.0113 (0.0153)
Expected bonus	-0.0164 (0.0100)
Expected bonus^2	8.45e-6 (6.87e-6)
Teacher male	-1.2526 (0.4654)
Teacher minority	0.2827 (0.2914)
Teacher certified	1.0453 (5.0335)
Teacher experienced	1.9652 (0.4184)
Class Size	-0.0899 (0.0182)
School size	0.0021 (0.0015)
Rural school	-1.1018 (0.2984)
Observation	10,046

†NCERDC 1995-1996 data. Estimation also included district and grade fixed effects.

Table 10: Effect of Incentives Using Pre-Incentive Data

Variable	Coefficient (Std. Dev.)
Number of Teachers	-0.0003 (0.0003)
Expected bonus	0.0004 (0.0003)
Expected bonus^2	2.60e-7 (2.17e-7)
Last year reading score	0.7906 (0.0024)
Male	-0.0679 (0.0041)
Minority	-0.1270 (0.0053)
Parent HS or less	-0.1302 (0.0045)
Teacher male	-0.0422 (0.0086)
Teacher minority	0.0068 (0.0074)
Teacher certified	-0.1057 (0.0372)
Teacher experienced	0.0284 (0.0086)
% Class minority	-0.0237 (0.0139)
% Class w/ Parent HS or less	-0.0013 (0.0115)
Class Size	-0.0036 (0.0009)
School size	4.13e-6 (3.18e-5)
Rural school	-0.0081 (0.0054)
Observation	84,766

†NCERDC 1995-1996 data. Estimation also included district and grade fixed effects.

Table 11: Classroom and Grade-level Targeting

	Status Quo	Classroom Targeting	Grade-level Targeting
Expected bonus	761.74 (81.48)	773.37 (263.61)	760.07 (116.69)
% Effort Increase		-4.18%	+3.16%
Δ Reading score†		-2.60	+1.74
% Teachers w/ higher effort		53.6%	90.5%
% Schools w/ higher scores		54.2%	97.0%
% Students w/ higher scores		52.8%	88.4%
Racial achievement gap‡	72.9	74.6	72.6
Among students/schools w/ higher scores			
Minority students	34.8%	33.4%	35.3%
Parents w/ HS or less	72.6%	70.6%	72.6%
School size	516.5	547.1	524.4
% Rural school	46.8%	44.6%	46.1%

†Change in percent of one standard deviation in test scores.
‡Percent of one standard deviation in test scores.

Table 12: Student and Teacher Outcomes from Tougher/Easier Bonus Standards

	Status Quo	10% Tougher	10% Easier	20% Easier	30% Easier	40% Easier
Expected bonus	761.74 (81.48)	704.04 (82.70)	816.78 (80.04)	871.96 (78.46)	926.45 (76.75)	980.32 (74.87)
% Effort Increase		-2.30%	1.46%	2.20%	1.99%	0.94%
Δ Reading score†		-1.30	0.81	1.16	1.02	0.51
% Teachers w/ higher effort		3.61%	87.6%	80.1%	69.2%	56.3%
% Schools w/ higher scores		3.41%	87.8%	79.5%	69.5%	56.3%
% Students w/ higher scores		3.84%	87.8%	80.2%	69.6%	55.9%
Racial achievement gap‡	72.9	73.2	72.6	72.4	72.2	72.1
Among students/schools w/ higher scores						
Minority students	34.8%	31.2%	35.8%	36.7%	37.9%	39.6%
Parents w/ HS or less	72.6%	65.1%	74.0%	74.6%	75.6%	77.3%
School size	516.5	426.1	519.4	517.3	515.1	510.6
% Rural school	46.8%	52.8%	46.5%	47.0%	47.1%	47.5%

†Change in percent of one standard deviation in test scores.
‡Percent of one standard deviation in test scores.

Figure 1: EOG Score Growth Rate Differences Before and After ABC

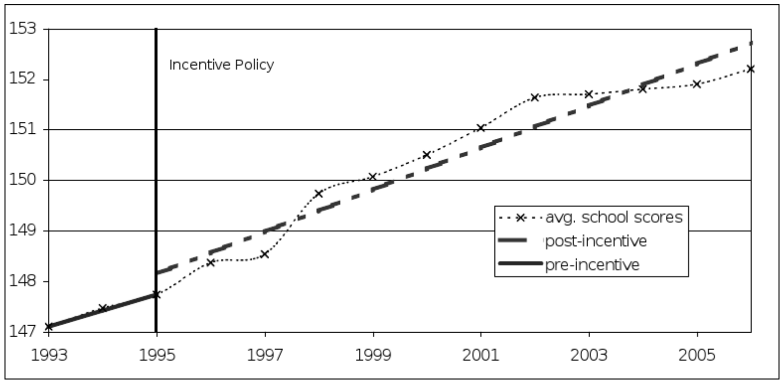

Figure 2: Non-cooperative Regime

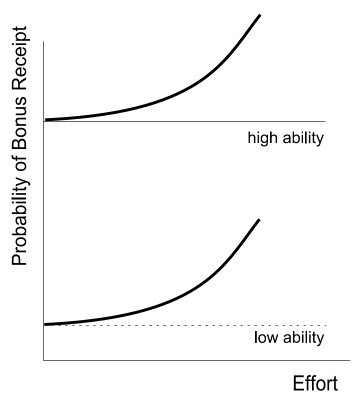

Figure 3: Cooperative Regime

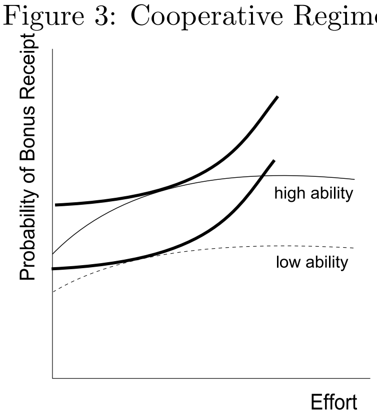

Figure 4: The effect of ϕ on effort exertion

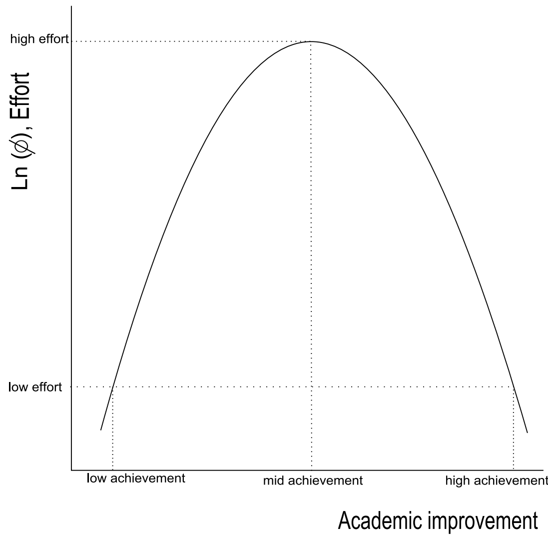

Figure 5: Illustrated Model

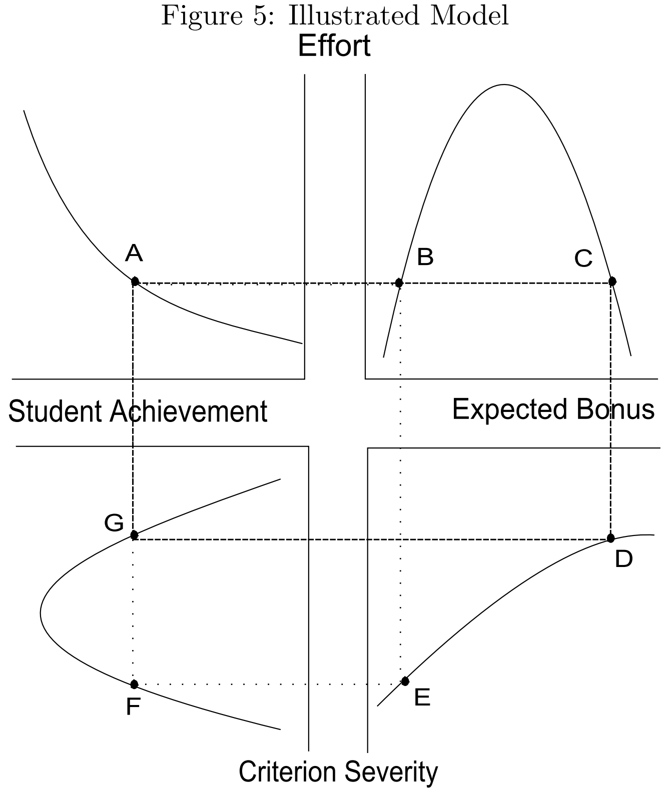

Figure 6: Histogram of Expected Bonus under Status Quo and Individual Criterion

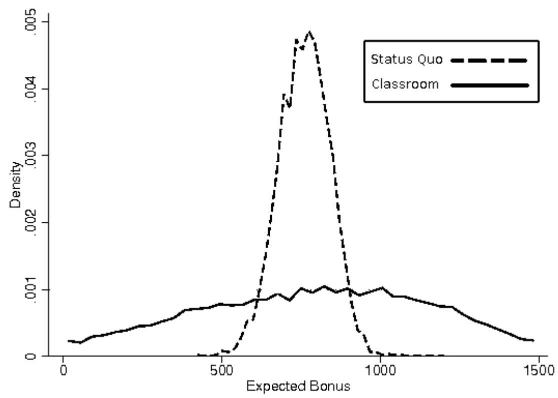

Figure 7: Histogram of Predicted Absence under Status Quo and Individual Criterion

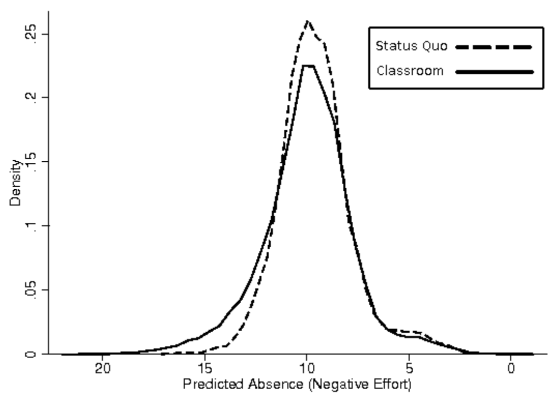

Figure 8: Histogram of Predicted Test Scores under Status Quo and Individual Criterion