Program on Education Policy and Governance Working Papers Series

The Impact of Incentives on Effort: Teacher Bonuses in North Carolina

Tom Ahn
University of Kentucky

Jake Vigdor
Duke University

PEPG 10-06

Prepared for the PEPG Conference
Merit Pay: Will It Work? Is It Politically Viable?

Harvard Kennedy School
Cambridge, Massachusetts
June 3-4, 2010

The Impact of Incentives on Effort: Teacher Bonuses in North Carolina*

Tom Ahn

University of Kentucky

Jake Vigdor

Duke University

Abstract

Teacher effort, a critical component of education production, has been largely ignored in the literature due to measurement difficulties. Using a principal-agent model, North Carolina public school data, and the state's unique accountability system that rewards teachers for school-level academic growth, we show that we can distill effort from teacher absence data and capture its effect on student achievement in a structural framework. We find that:
1. Incentives lead teachers to try harder. The bonus program reduced the number of sick days taken by about 0.6 days for an average teacher.
2. When teachers try harder, students do better. Increased effort of teachers translates into improved student performance. Estimates show that standardized reading scores increased by about 1.3% of a standard deviation and standardized math scores by about 0.9% of a standard deviation.
3. Group-level incentives can actually be more powerful than individual-level incentives. Policy simulations from the model estimates show that an individual bonus program would actually produce weaker incentive effects. While free-rider effects are eliminated, individual incentives push a majority of teachers into one of two categories: those who would qualify for the bonus even without trying and others would not qualify no matter how hard they worked.

Performance pay for teachers: are school-level incentives enough?

Over the past ten years, researchers have devoted considerable effort to the measurement of the output of schools and teachers, using standardized test scores.[1] Our ability to infer the quality of teaching in a school or classroom has developed sufficiently far that school

* We grateful to the Center for Child and Family Policy for access to the NCDPI dataset and to the American Enterprise Institute for financial support.

[1] See Rivkin, Hanushek, and Kain (2005), Clotfelter, Ladd,and Vigdor (2007), Goldhaber and Anthony (2007), and Rockoff (2004), among many others.

districts across the nation, from Denver to Washington DC and many points in between, have put incentive programs in place that make student test score performance a major factor in the evaluation, and in some cases the compensation, of teachers.

There are some sticky issues, however, about how to really make a pay-for-performance scheme based on test scores operate.[2] For one thing, most school systems don't conduct standardized tests in every grade, or in every subject. How exactly are we supposed to measure the performance of a kindergarten teacher? Or a high school Spanish teacher? Or a middle school physical education teacher? And if we don't evaluate these teachers' performances, how do we pay them?

A second issue with pay-for-performance schemes is that they may lead teachers to fight amongst each other for the best students. Teachers might perceive that certain students hit performance targets more easily; in many cases there is strong evidence to back up this perception. Principals and other administrators might come under pressure to fiddle with classroom assignments. In theory, a commitment to randomly assign students to classrooms would present teachers with a level playing field. But even randomization makes some people lucky and others unlucky sometimes. To complicate matters, many schools have instituted different tracks for students with differing academic abilities. Pay-for-performance would add a financial reason to lobby to teach honors-class students in addition to the laundry list of non-pecuniary benefits.

[2] See Neal (2009) for a review of some of the inefficiencies of the pay-for-performance system.

Another problem is the “noisy" test results issue. The “noise” we are talking about here is statistical noise, and it is a more severe problem for classroom-sized groups of students relative to school-sized groups. A “noisier” test result makes it difficult to discern whether a higher average score for a group of students relative to another group really means that the former group of students knows more about the subject they were tested on. The smaller the number of test takers, the higher the possibility that one outlying result can throw off the entire result. For instance, if one or two students were ill on the day they took the test, their poor exam scores will significantly drag down the class average and make the teacher look less competent than she really is. These unfortunate students would have a smaller impact on the school average, since they make a smaller proportion of the larger student body.

These three Gordian knots – the presence of untested grades and subjects, non-random assignment of students to teachers, and the statistical noise problem in small samples – could be sliced with one modification to pay-for-performance: reward teachers on the basis of all students in the school, rather than just those in their classroom. With school-based incentives, we need not worry about what to do with teachers of odd subjects, or in untested grades. And we need not worry about teachers fighting one another in a zero-sum game, or about statistical noise leading to good teachers going unrewarded.

The primary theoretical argument against school-based rewards will be familiar to economics 101 students everywhere: the free-rider problem. Compared with an individual-level incentive, a group-level incentive should have less impact. When your

own effort determines my compensation, you have a very strong reason to work hard. When the combined effort of a large group determines your compensation, you may feel at greater liberty to slack off, since most of your reward depends on the actions of other people anyway. It's the tragedy of the commons, the prisoner's dilemma – whatever you want to call it. It's an argument so strong and so intuitive that you don't hear many education economists saying that school-level rewards are the way to go.

Until now. New evidence, derived from the experiences of North Carolina public schools, which have implemented school-based monetary incentives for more than a decade now, indicates that this conventional wisdom – that individual incentives are more powerful than group incentives – is in fact wrong. Yes, Virginia, there is a free rider effect. But what the conventional wisdom fails to incorporate is a powerful countervailing effect, which we might call the "tortoise and hare" effect, borrowing from Aesop's fables. Consider the following scenario. You are an excellent teacher – one of the best in the business. If the school system sets a bar and promises you rewards if your students exceed it, you know you can exceed the expectation even without trying. Like the hare in the fable, your incentive to try your best is undermined by a sense that your success is inevitable. We may fault the hare for his laziness, but is this really such a surprising response when victory seems assured?

The teacher next door, on the other hand, is hopelessly incompetent. You know as well as she does that no matter where the bar is set, her students will almost certainly fall below it. Like the tortoise in the fable, it is only her personal virtue that implores her to

exert effort: the incentive means very little. Again, we may cheer the tortoise for his perseverance, but how wise is it to expend effort when there is virtually no chance of success? So, for both you and your neighbor, the individual-level incentive scheme provides almost no incentive to exert greater effort. You are bound to be rewarded no matter what, and your neighbor is destined to fail no matter what. It would be great if both the tortoise and the hare tried their hardest regardless of the competition, but the most likely outcome is that the hare would win walking backwards, and the tortoise would quit before the race even begins.

Now suppose we tie you and your neighbor together: your reward will be based not on what you do individually, but the sum total of what you accomplish. All of a sudden, you recognize that the status of your reward is in doubt, and the teacher next door realizes that she now has a realistic shot at the reward; both you and your neighbor are going to have to exert some effort to ensure that the average across your two classrooms exceeds the standard.

While the traditional moral of the fable is that "slow and steady wins the race," perhaps we should reconsider the wisdom of such a match-up in the first place. Rather than race tortoises against hares, we should pair one of each together and judge each pair by their combined time. In this scenario, each competitor faces a stronger incentive to excel, because it is their team's average time that matters, not their rank within the team.

This is a plausible scenario, right? But we just don't know how common this scenario might be. How often do very good and very poor teachers share the same school? And just how powerful is this free rider effect anyway? The answers lie in the research. To tell you about the research, we first need to spend some time getting to know the setting.

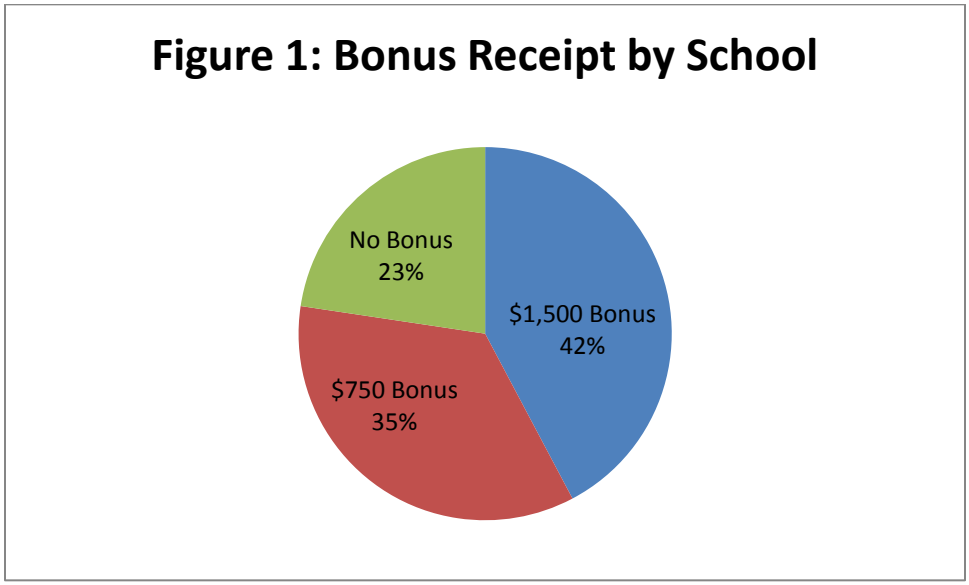
The North Carolina State Accountability System

The North Carolina ABC accountability program (ABC is an acronym for Accountability, teaching the Basics, and emphasis on local Control) began in the 1996/97 school year. In its inaugural year, teachers in elementary and middle schools were awarded a cash bonus of $1,000 if the school's average year-over-year improvement in reading and math test scores exceeded the required threshold set by the state. In the following year, the bonus program was extended to high schools, and the award became two-tiered, with teachers receiving $750 in schools that cleared a first threshold referred to as "expected" growth in test scores and $1,500 in schools that cleared a more stringent "exemplary" or "high" growth threshold. [3]

Education authorities face a delicate balancing act in setting criteria for bonus payments. If teachers perceive that there is no chance of receiving a bonus, or conversely that the bonus is a sure thing, they have little reason to alter their behavior. This is a basic statement of the "tortoise and hare" effect described above. Fortunately, in North Carolina's case, teachers in most schools face real uncertainty about the amount of their

[3] A complete description of the bonus program and the formulas used to set each school's threshold can be found in Vigdor (2008).

bonus. Figure 1 shows the proportion of schools in the 1999/00 to 2001/02 school years qualifying for $750 or $1500 bonuses. Roughly three-quarters of the schools in the state received bonus payments, but less than half received the full $1,500. The average bonus paid out is roughly $890 (0.23 X $0 + 0.35 X $750 + 0.42 X $1,500 = $890). Vigdor (2009) presents additional evidence that among the schools eligible for any bonus at all, about half receive the full $1,500. There are very few schools that can count on the full $1,500 as a sure thing, and very few for which the $750 standard is completely unattainable.



Incidentally, North Carolina's system is made possible because the state has a longitudinal data system that can link the performance of individual students as they progress from grade 3 to grade 8. Many other states, unfortunately, have no capacity to link students across years, implying that they can only judge schools by how the students perform in a given year, not by how much they improve in a given year. This limitation

forced the federal No Child Left Behind act to focus on proficiency rather than improvement. Why does this matter? A school that serves very low-performing kids and manages to improve their performance dramatically might not be rewarded if their ultimate performance is below the state's threshold for proficiency.

Figuring out what the bonus program accomplishes

The North Carolina ABC system is not costless. Their state legislature needs to allocate 90 million dollars or more per year for these performance bonuses. And while there's a strong economic reason for thinking that performance bonuses improve student performance, there's no specific guidance regarding how big the impact should be, let alone whether the impact is worth the amount of money being spent on the program.[4]

So is the program worthwhile? How can we tell? The gold-standard method of evaluating a program such as the ABC initiative would have been to conduct a randomized trial. Schools in North Carolina would have been randomly assigned into two groups: a "treatment group" of schools where teachers were awarded the bonus according to the ABC framework and a "control group" where teachers did not receive the bonus. If the incentives worked as planned, teachers in the treatment group would have exerted higher effort to teach students, and this would have translated into higher scores for students in treatment schools relative to those in control schools.

[4] See Figlio and Kenny (2007).

In North Carolina, all public schools became eligible for the bonus at the same time. This greatly complicates any effort to evaluate the program. The best feasible method to study the effect of merit pay would be to look at a "before" and "after" ABC implementation snapshot of student performance. If the distribution of students and teachers and characteristics of schools remained constant over time, we could compare the performance of students before and after the teachers started receiving bonuses to see if the money led to increased academic achievement. Unfortunately, during the 1990's and 2000's, North Carolina experienced large population changes. The state's population increased by more than 20% between 1990 and 2007. Its demographic makeup has changed as well. For example, the state's Hispanic population exploded between 1990 and 2007; the group formed 1.2% of the population in 1990 and 6.7% in 2007. These changes in the underlying composition of the population, plus other alterations to educational practice, probably would have led to a change in achievement levels even in the absence of the bonus program. It isn't possible to distinguish what trends are attributable to the bonus program and which to these confounding trends.

Solving the evaluation problem

There is one potential way out of this conundrum, and it involves taking advantage of the "free rider" problem we spoke of earlier. Whatever the impact of the bonus program, be it positive, negative, or nil, we would expect a stronger impact in smaller schools, given the nature of the group-level incentive. By "smaller school," what we mean is schools with fewer teachers. In a one-room schoolhouse, one person's effort is all that counts,

and we'd expect incentives to have a very strong impact. In a monolithic urban school, the group-level incentives should have a weak effect on individual teachers. So, in the wake of the bonus program, we expect differences to open up between smaller and larger schools. If the performance of students in small schools accelerated relative to students in larger schools after the bonus program began, the bonus program is the most likely explanation. If, on the other hand, there was no differential trend across schools of different sizes, the logical conclusion is that the bonus program had little impact.

There's a second, related avenue to consider. Not every school stands the same chance of meeting the bonus criterion. Teachers know this as well as everybody else. In highly dysfunctional schools, there is little chance that teachers will raise student performance sufficiently to merit a monetary reward. So why bother? At the other end of the spectrum, teachers in privileged schools might recognize that their students will meet expectations even if they turn in a mediocre effort. So once again, why bother? It's only in the schools that are truly on the margin where effort matters. So we expect the greatest improvements in those schools where the likelihood of receiving the bonus is truly in doubt. We can infer which schools those are on the basis of past performance, or on the more basic characteristics of the students themselves.

So, we expect the performance of small schools, and in schools on the margin for receiving a bonus, to improve relative to others. We expect teachers in both smaller and marginal schools to exert greater effort, and we further expect this to translate into academic improvements for students. One could eliminate effort from the equation and

just look for patterns in test scores. This strategy has problems, however, if test scores are the result of more than just teacher effort.

Education researchers have documented many ways in which incentive programs have unintended consequences. For instance, there has always been the fear that teachers will "teach to the test," resulting in better test scores, but not more learning. Other, more underhanded methods have also been documented. Principals have been observed classifying marginal students as disabled or suspending them immediately before an exam date, resulting in fewer of these students, who are expected to perform poorly, being counted. Teachers in some instances have been known to change the answer sheets filled out by students to fabricate a higher score. Schools have also been shown to up the calorie content of meals on the day of the exam. All schools have incentives to engage in this kind of behavior, regardless of their size.[5] Any one of these behaviors is problematic from a policy perspective, because they imply that schools have found ways to manufacture higher test scores without providing a better education. So, ideally, we'd like to verify that the incentive scheme has an impact on a factor that really correlates with better student learning.

We can make some inferences about how hard teachers work by observing how often they call in sick during the school year. When a teacher has an unscheduled absence during the school year, she is doing it with the knowledge that it will be detrimental to student learning. A substitute teacher will have to be assigned and lesson plans will be

[5] See Cullen and Reback (2002), Jacob (2005), Grissmer and Flanagan (1998), Hanushek and Raymond (2005), among many others.

thrown off track. Several studies have shown that students learn less in years when their teacher takes more absences.[6] This basic pattern might reflect either the low quality of substitute teachers, or the negative impacts of having a teacher who is less motivated to come to work in the morning. These teachers might take fewer absences, but they also might exert less effort in many other ways.

One key insight here is that teacher absences are a signal of an underlying, and more important factor: what social scientists would term a "latent variable." The latent variable in this case is something that increases when there is a bonus at stake, and causes teachers to take fewer absences. We'll call this latent variable effort.[7]

Our basic prediction, then, is that the ABC Bonus program would have created a "teacher absence gap" between small schools, where the teachers were strongly incentivized, and larger schools where the incentives had a smaller impact. At the same time, schools in the middle of the pack should have improved relative to those at either end. Only the data can tell us, though, exactly how large these effects might be.

Inferring the potential impact of individual-level incentives

How exactly can anything informative be said about individual-level incentives, in a state where only group-level bonuses have been implemented? To be sure, there is some extrapolation involved, but it's a modest stretch. This analysis will tell us what sort of

[6] See Clotfelter, Ladd, and Vigdor (2007), and Ehrenberg et al. (1991).
[7] Note that we abstract away from student effort. See Angirst and Lavy (2002) for a discussion.

impact the bonus program will have, as a function of school size and the likelihood of hitting the benchmark. It's easy to contemplate an individual-level incentive scheme as a variant of this. Just imagine that teachers work alone, and have likelihoods of receiving a bonus tied to the performance of their own students, rather than the school as a whole. Using the results of the exercise, we can easily predict the likely impact on teacher effort and student performance.

The evidence: why you should stop worrying about group-level incentives

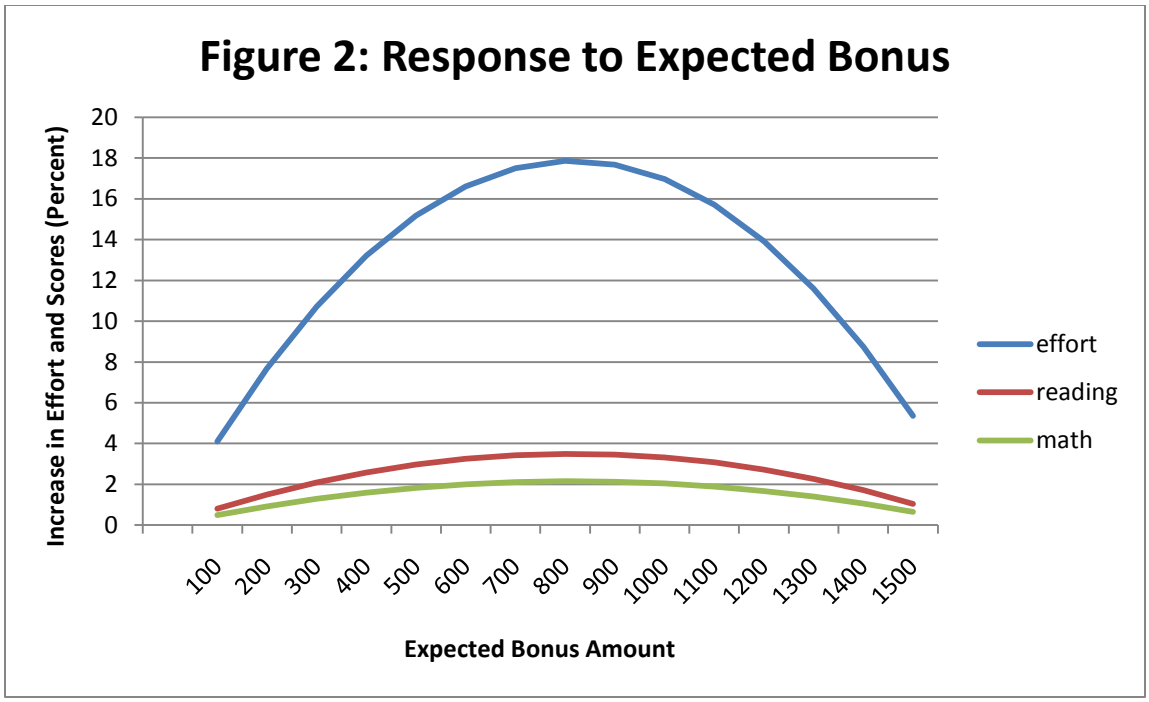
We have a way of measuring free-rider effects and teacher effort, so we next need to see if the bonus incentives really work as planned. Are teachers motivated by money? It's naïve to assume that teachers are *not* motivated by money, as naïve as assuming that teacher are *only* motivated by money.[8] The relevant question is: can teachers be motivated to give more effort compared to the status quo at reasonable cost? The answer is yes. Comparing a teacher's absenteeism rate when school is in session and the expected dollar amount of the bonus she is expected to receive, we find that an increase in likelihood of qualifying for the bonus will cause her to take fewer absences. If we were to take an average teacher who has a very small chance at qualifying for the bonus (where her expected bonus is equivalent to $400) and increased her probability of qualifying for the bonus (so that her expected bonus becomes $900) we expect her to take about one fewer sick day over the course of a school year. In terms of the underlying effort variable, the incentive effect of the extra $500 at stake is a 10% boost to effort.

[8] See Carnoy and Loeb (2002), Jacob (2007), and Vigdor (2008) for examples of teacher response to outside pressures or enticements.

While this seems like a rather cost-effective way to improve teacher performance, remember that the strength of incentives is highly sensitive to the perceived likelihood of receiving a bonus. Imagine how motivated a teacher would be to put in extra effort if the likelihood of qualifying for the bonus was 100%. As the graph below shows, policy makers should take care not to make the bonus too easy or too difficult to get, as either extremes will do little to motivate teachers.

Of course, increased effort is nice, but this only matters if it actually translates into more learning. So do students learn more with motivated teachers? Again, the answer is yes. A highly motivated teacher will raise her student's standardized test scores by a significant amount. An average teacher who is efficiently motivated in the current NC incentive system is expected to raise her students' average reading scores by more than 3.5% of a standard deviation, and math scores by about 2.2% of a standard deviation. For an elementary school teacher of 20 students, the bonus program spends an average of $6.25 to raise the performance of one student in one subject by 1% of a standard deviation. This implies that incentive programs such as North Carolina's are far more cost-effective than other popular education interventions, such as reducing class sizes.[9]

[9] See Mulralidharan and Sundararaman (2009) for an evolution of teacher pay-for-performance in a foreign setting.



With statistical evidence that teachers are motivated to work harder by cash rewards, and that motivated teachers get students to perform better on exams, we can now address the question of whether we should push hard for individual incentives. The current system in North Carolina treats the school as one unit. That is, the threshold that the teacher must surpass in order to receive the bonus depends, not just on her students, but on the performance of all tested students at the school. How would students fare if the state bonus program rewarded individual performance?

The advantages of a purely individual-level incentives system seem self-evident. If we are going to spend extra public funds to get teachers to do their jobs better, we should, at least, make sure that every dollar we spend will have the most bang for our buck. Our intuition from economics tells us that we can get rid of inefficiencies from free-rider effects by evaluating bonus at the individual level. This seems to be a powerful case for

going to bat for individual-level incentives. Or is it? As we will see, the answer is not so simple. School-level incentives may have been instituted for political and administrative expediency, but could it be that the state has actually backed into a more effective system than individual-level incentives?

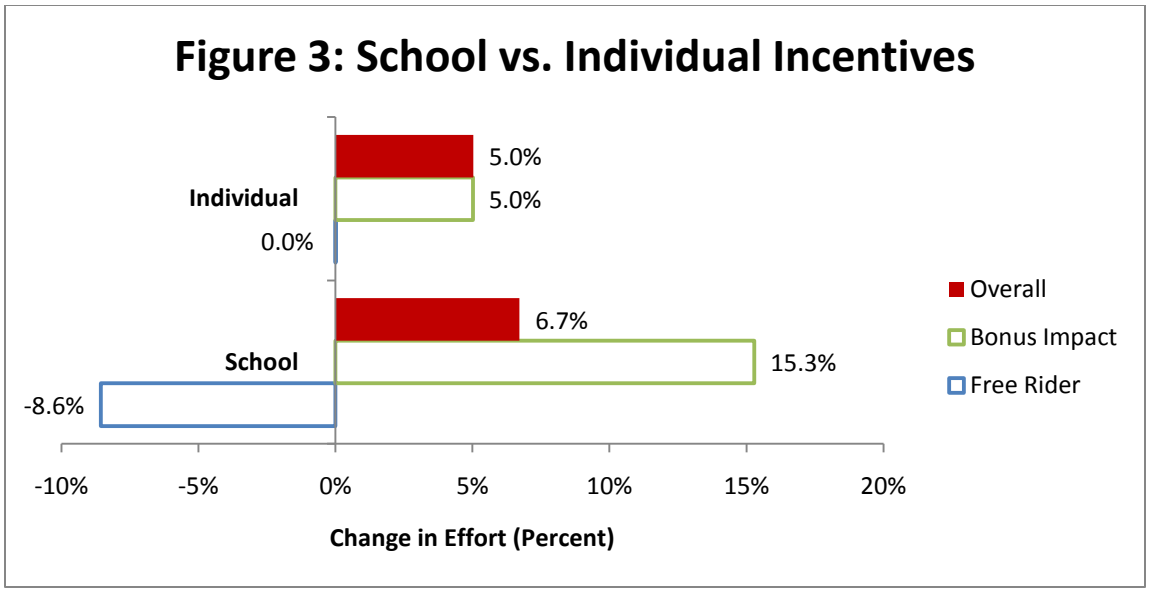
As we mentioned in the introduction, the key argument for school-level incentives is that tying low-ability and high-ability teachers together may force both groups to exert higher effort to qualify for the bonus. The lack of initial motivation for both teachers is that they are both too far away from the bar set by the government. The high-ability teacher is too far above the bar, allowing her to coast yet still qualify for the bonus, and the low ability-teacher is too far below the bar, effectively preventing her from receiving the cash, no matter how hard she tries. The insight is that the teacher's motivation decreases as the distance from the bar increases in *either direction*.

It isn't difficult to see that low-ability teachers will be discouraged when the bar becomes too difficult to reach and that high-ability teachers will become complacent when the bar becomes too easy to reach: this is precisely the hare-and-tortoise effect. The advantage of a school-level incentive is that it can simultaneously lower the bar for low-ability teachers and raise the bar for high-ability teachers. Because high ability teachers are tied together with low ability teachers, their average score declines. While previously coasting to the bonus, they will now have to pull that much harder to make it over the bar, all the while dragging the extra weight created by low ability teachers. Low ability teachers, on the other hand, see their average scores increase. With the boost from high ability

teachers, they have a decent chance at qualifying for the bonus, if they put in extra effort. This induces both groups of teachers to try harder.[10]

So, which is better? School or individual incentives? In North Carolina, it appears that changing from the school to individual incentives would not yield the widely predicted increase in teacher effort and student achievement. As the system is converted from school-level incentives to individual incentives, free-rider effect is eliminated. At the same time, the change introduces the hare-and-tortoise effect by pushing most teachers away from the state standard. These two effects pull in opposite directions. Like a tug-of-war or a see-saw, it is impossible to decrease one effect without increasing the other. We find that the latter effect dominates the former, and average teacher effort, and therefore average student achievement, declines in the individual incentive regime relative to the group incentive regime (See Figure 3). Consider an average-sized NC elementary school with about 35 full time teachers. In the group incentive regime, teachers who ignored the free rider effect would expand their effort by 15%. The free-rider effect saps more than half of this expansion, leading teachers to exert just 6.7% more effort. The individual incentive regime eliminates the free-rider effect; but because a higher proportion of teachers view bonus receipt as either a sure thing or an unattainable goal, the average impact on effort is in fact lower.

[10] See Booher-Jennings (2005) and Neal and Schanzenbach (2009) for other examples of the distributional impacts of accountability.



Both school and individual incentives result in increases in teacher effort and subsequent gains in test scores. The gains under the school incentive are larger than the gains under individual incentives, because the larger increase in effort due to the hare-and-tortoise effect more than offsets the loss in effort due to free-rider effects.

To be fair, there is one potential method of reducing the hare-and-tortoise effect. Rather than implement an all-or-nothing bonus, one could offer teachers a continuously varying performance-based salary supplement. Each incremental gain in student achievement would be associated with an incremental increase in teacher pay. The problem with such a scheme, of course, is that it magnifies the various problems associated with individual-level schemes as outlined above. Continuously varying bonuses would reward teachers for statistical flukes, and could never realistically be implemented for teachers in untested grades or subjects. The incentive effect of a large dollar amount, awarded when scores

pass a distinct threshold, might also be quite a bit stronger than the promise of just a few dollars for a marginal improvement.

Conclusion

The economic rationale for incentivizing teachers is strong, but efforts to implement pay-for-performance plans have often foundered on the details. Pay-for-performance is hard to apply to teachers in untested grades, or in untested subjects. Individual-level schemes threaten to introduce wasteful competition among teachers for the best students. And concerns about the statistical reliability of test scores implies that education authorities might have to wait for years before rewarding deserving teachers, dismissing ineffective ones, or devoting attention to those who could really excel with a little bit of help.

The headlong rush to individual-level incentive schemes has occurred under the presumption that free-rider effects would hobble school-level incentives. Since the average test score at the school is largely out of the control of individual teachers, the argument goes, the bonus does not serve as a strong incentive.

In fact, the cost-effectiveness of a well-designed group-level incentive can be significantly better than an equivalently constructed individual-level incentive. Moving to individual incentives increases each teacher's distance from the bar, introducing a hare-and-tortoise effect more severe than the free-rider effect.

This analysis also verifies a point that should fall within the realm of common sense, but bears repeating here: incentives don't really accomplish anything if they are impossible to obtain, or if they are impossible *not* to obtain. The power of incentives arises in scenarios when individuals realize that something of value is at stake. There will always be pressure to water down incentive schemes to the point where they serve as nothing more than a guaranteed pay raise. Those who wish to implement pay-for-performance must be prepared to resist this pressure.

North Carolina's experience verifies that teacher incentives can improve student performance, even in the presence of the dreaded free-rider effect. If the policy argument comes down to a choice between a consensus on school-level incentives and a protracted fight over individual-level incentives, proponents of pay-for-performance should save their ammunition for other battles.

Bibliography

[1] Angrist, J. and V. Lavy (2002) "The Effect of High School Matriculation Awards: Evidence from Randomized Trials" NBER Working Paper No. 9389.

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

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

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

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

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

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

[8] Figlio, D., L. Kenny (2007) "Individual teacher incentives and student performance," Journal of Public Economics, Elsevier, vol. 91(5-6), 901-914.

[9] Goldhaber, D., E. Anthony (2007) Can Teacher Quality be Effectively Assessed? National Board Certification as a Signal of Effective Teaching. Review of Economics and Statistics, 89(1): 134-150.

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

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

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

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

[14] Muralidharan K., V. Sundararaman (2009) “Teacher Performance Pay: Experimental Evidence from India." NBER Working Paper 15323.

[15] Neal D., (2009) “Designing Incentive Systems for Educators" forthcoming in Performance Incentives: Their Growing Impact on American K-12 Education, edited by Matthew Springer, Brookings.

[16] Neal D., D. Schanzenbach (2009) “Left Behind by Design: Proficiency Counts and Test-Based Accountability." Review of Economics and Statistics, Forthcoming.

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

[18] Rockoff, J. (2004) “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." American Economic Review 94(2): 247-52.

[19] Rothstein J. (2008) “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." Quarterly Journal of Economics, Forthcoming.

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