Education Outlook

Making Teacher Incentives Work: Lessons from North Carolina's Teacher Bonus Program

By Thomas Ahn and Jacob L. Vigdor

North Carolina has operated one of the country's largest pay-for-performance teacher-bonus programs since the late 1990s. New research shows that a North Carolina–style incentive-pay program has the potential to improve student learning by encouraging teachers to exert more effort on the job. The North Carolina model avoids three pitfalls associated with implementing individual-level pay-for-performance plans: the problem of grades and subjects without standardized tests, the problem of teachers fighting for the best students, and statistical noise in test scores. The program also enjoys broad political support, including from the state teachers union. Education reformers worldwide should understand how performance pay can improve student learning.

Over the past ten years, researchers have devoted considerable effort to measuring the output of schools and teachers using standardized test scores. The ability to infer teaching quality in a school or classroom has developed enough that school districts across the country have put incentive programs in place that make student test-score performance a major factor in teacher evaluation, and sometimes compensation.

The economic rationale for incentive programs is strong. Typical teacher contracts reward credentials, such as years of experience and postgraduate degrees, even though they often have no proven association with improved student learning. After a brief probationary period, teachers can expect the same compensation and career path regardless of their effort to improve student achievement. Success in teaching brings personal satisfaction, but little else. For the most highly motivated teachers, this may be enough, but for some the absence of a direct reward inhibits hard work.

Thomas Ahn (thomas.ahn@uky.edu) is an assistant professor of economics at the University of Kentucky. Jacob L. Vigdor (jacob.vigdor@duke.edu) is a professor of public policy and economics at Duke University.

Key points in this *Outlook*:

- North Carolina teachers receive pay supplements of up to $1,500 when the standardized test performance of all students in their school improves by more than a predetermined amount.

- The bonus program leads teachers to exert more effort on the job: the average teacher took 0.6 fewer sick days, and standardized test scores rose by about 1.3 percent of a standard deviation in reading and 0.9 percent in math.

- Results indicate that individual-level incentives would actually have a weaker effect than school-level ones. Many teachers would qualify for individual-level bonuses without trying, and others would not qualify no matter how hard they tried.

- Teacher incentives are cost-effective. Compared to other popular education reforms, such as reduced class sizes, incentives provide more than four times the amount of student improvement per dollar spent.

Incentive programs promise to restore this reward mechanism, which is the hallmark of most private-sector occupations. The objective measurement of student success, made possible by standardized testing, appears to be the final ingredient necessary to make performance incentives a reality.

There are some sticky issues, however, in operating a pay-for-performance scheme based on test scores. For one thing, most school systems do not conduct standardized tests in every grade or subject. How can we measure the performance of a kindergarten teacher, a high school Spanish teacher, or a middle school physical education teacher? And if we do not evaluate these teachers' performances, how do we pay them?

Not many education economists advocated school-level rewards. Until now.

A second issue with pay-for-performance schemes is that they may lead teachers to fight among each other for the best students. Even if performance measures are based on "value added," implying that teachers are not rewarded simply for having students who were above average at the beginning of the year, teachers might perceive that certain students hit performance targets more easily; in many cases, there is strong evidence to back up this perception.[1] Principals and other administrators might come under pressure to fiddle with classroom assignments. In theory, randomly assigning students to classrooms would present teachers with a level playing field, but even randomization sometimes makes some people lucky and others unlucky. To complicate matters, many schools have instituted different tracks for students with different academic abilities. Pay-for-performance would give teachers a financial incentive to lobby to teach honors-class students in addition to the laundry list of nonpecuniary benefits.

A third problem is the "noisy" test results issue. A noisier test result makes it difficult to discern whether a higher average score for a group of students relative to another group means that the former group knows more about the subject than the latter. Statistical noise is a more severe problem for classroom-sized groups of students than for school-sized groups. The fewer the test takers, the higher the possibility that one outlier can throw off the entire result. For instance, if one or two students were ill when they took a test, their poor exam

scores could significantly drag down the class average and make the teacher look less competent. These unfortunate students would have a smaller impact on the school average, however, since they make up a smaller proportion of the larger student body.

These three Gordian knots—untested grades and subjects, nonrandom 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 classrooms. With school-based incentives, we need not worry about what to do with teachers of untested subjects and grades, or teachers fighting one another in a zero-sum game, or statistical noise leading to good teachers going unrewarded.

The primary theoretical argument against school-based rewards is familiar to economics 101 students: the free-rider problem. Compared with an individual-level incentive, a group-level incentive usually has less impact. When effort determines compensation, individuals have a strong reason to work hard. When the combined effort of a large group determines compensation, however, they may feel at liberty to slack off, since most of the reward depends on the actions of other people. It is the tragedy of the commons, or the prisoner's dilemma, and it is an argument so strong and intuitive that not many education economists advocated school-level rewards.

Until now. New evidence from North Carolina public schools, which have implemented school-based monetary incentives for more than a decade, indicates that this conventional wisdom—that individual incentives are more powerful than group incentives—is wrong. Yes, there is a free-rider effect, but the conventional wisdom fails to incorporate a powerful countervailing effect, what could be called the "tortoise and hare" effect, borrowing from Aesop's fables. Consider the following scenario. One teacher is excellent—one of the best in the business. If the school system sets a bar and promises her rewards if her students exceed it, she knows she can exceed the expectation without trying. Like the hare in the fable, her incentive to try her best is undermined by a sense that success is inevitable. We may fault the hare for his laziness, but is this really such a surprising response?

The teacher next door, however, is hopelessly incompetent. Everyone knows 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 few of us would persist so doggedly in the face of overwhelming odds. For both teachers, the individual-level incentive scheme provides almost no incentive to exert greater effort. One is bound to be rewarded, and the other is destined to fail. For both tortoise and hare, with little doubt about the outcome, the only incentive to exert effort is a sense of personal satisfaction. Thus, the individual-incentive regime threatens to end up a lot like the status quo.
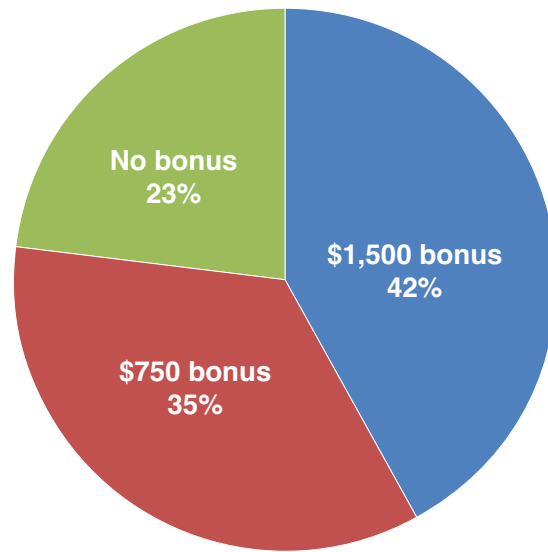
Now suppose both teachers are tied together: their reward will be based not on what they do individually, but on the sum total of what they accomplish. Suddenly the excellent teacher recognizes that the status of her reward is in doubt, and the teacher next door realizes that she now has a realistic shot at the reward. Both are going to have to exert some effort to ensure that the average across their 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 matchup in the first place. Rather than pit tortoises against hares in an unfair competition, we could assemble teams of mixed tortoises and hares and judge each team by its combined performance. In this scenario, each competitor faces a stronger incentive to excel because her team's average time matters, not her rank within the team. We do not know how common this scenario might be. How often do very good and very poor teachers share the same school? And how powerful is the free-rider effect? The answers lie in the research. But first we need to understand the accountability program.

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 threshold set by the state. In the following year, the bonus program was extended to high schools, and the award became

SOURCE: Authors' calculations.

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.[2]

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 that the bonus is a sure thing, they have little reason to alter their behavior—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 bonus. Figure 1 shows the proportion of schools in the 1999–2000 to 2001–2002 school years qualifying for $750 or $1,500 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 was roughly $890 (0.23 x $0 + 0.35 x $750 + 0.42 x $1,500 = $892.50). Evidence shows that among the schools eligible for a bonus, about half receive the full $1,500.[3] There are few schools that can count on the full $1,500 as a sure thing, and 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 links the performance of individual students as they progress from grade three to grade eight. Many other states, unfortunately, have no capability to link students across years, implying that they can only judge schools

by how students perform in a given year, not by how much they improve. 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 students and manages to improve their performance dramatically might not be rewarded if their ultimate performance is below the state's threshold for proficiency.

What Does the Bonus Program Accomplish?

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

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 to 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.

In North Carolina, all public schools became eligible for the bonus at the same time, which greatly complicates efforts to evaluate the program. The best feasible method to study the effect of merit pay is to look at a snapshot of student performance before and after ABC implementation. 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 1990s and 2000s, North Carolina experienced large population changes. The state's population increased by more than 20 percent between 1990 and 2007. Its demographic makeup changed as well. For example, the state's Hispanic population exploded between 1990 and 2007; Hispanics formed 1.2 percent of the population in 1990 and 6.7 percent 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 is not possible to distinguish which trends are attributable to the bonus program and which to these confounding changes.

Solving the Evaluation Problem. There is one potential way out of this conundrum, and it involves taking advantage of the free-rider problem described earlier. Whatever the impact of the bonus program—positive, negative, or nil—we would expect a stronger impact in smaller schools, given the nature of the group-level incentive. By "smaller schools," we mean those with fewer teachers. In a one-room schoolhouse, one person's effort is all that counts, and we expect that incentives would have a very strong impact. In a monolithic urban school, however, 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 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.

Policymakers should take care not to make the bonus too easy or too difficult to get, as either extreme will do little to motivate teachers.

There is a second, related avenue to consider. Not every school stands the same chance of meeting the bonus criterion. Teachers know this as well as anybody. 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 recognize that their students will probably meet expectations even if they turn in a mediocre effort. So once again, why bother? Only in schools on the margin does effort matter. We expect the greatest improvements in schools where the likelihood of receiving a bonus is in doubt, and we can infer which schools those are based on past performance or the students' basic characteristics.

We also expect the performance of small schools and schools on the margin to improve relative to others. Teachers in smaller and marginal schools will likely exert greater effort, and we 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 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, so fewer poor-performing students are tested. Teachers in some instances have changed students' answer sheets to fabricate a higher score. Schools have also increased 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. These behaviors are problematic from a policy perspective because they imply that schools have found ways to manufacture higher test scores without providing a better education. Ideally, we would verify that the incentive scheme has an impact on a factor that correlates with better student learning.

The Underlying Factor: Effort. Some inferences can be made about how hard teachers work by observing how often they call in sick during the school year. When a teacher has an unscheduled absence, she knows that it will be detrimental to student learning. A substitute teacher will have to be assigned and lesson plans will be thrown off track. Several studies have shown that students learn less in years when their teacher takes more absences. This basic pattern might reflect either the low quality of substitute teachers or the negative impact of having a teacher who is less motivated to come to work. These teachers might take fewer absences, but they also might exert less effort in many other ways.

One key insight 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 will call this latent variable "effort."

Our basic prediction is that the ABC bonus program will create a "teacher absence gap" between small schools, where the teachers have strong incentives, and larger schools, where the incentives have a smaller impact. At the same time, schools in the middle of the pack should improve relative to those at either end. Only the data can tell us exactly how large these effects will be.

How 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 is a modest stretch. Our analysis shows the impact the bonus program will have as a function of school size and the likelihood of hitting the benchmark. It is 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: Stop Worrying about Group-Level Incentives

We have a way of measuring free-rider effects and teacher effort. Next we need to see if the bonus incentives work as planned. Are teachers motivated by money? It is just as naïve to assume that teachers are *not* motivated by money as it is to assume that they are *only* motivated by money. The basic question is whether teachers can be motivated to give more effort compared to the status quo at a reasonable cost. The answer is yes. Comparing a teacher's absenteeism rate when school is in session and the dollar amount of the bonus she expects to receive, we find that an increase in the likelihood of qualifying for the bonus will cause her to take fewer absences. If we take an average teacher who has a very small chance of qualifying for a bonus (where her expected bonus is equivalent to $400) and increase her probability of qualifying for the bonus (so 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 percent boost to effort.

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 were 100 percent. As figure 2 shows, policymakers should take care not to make the

bonus too easy or too difficult to get, as either extreme will do little to motivate teachers.
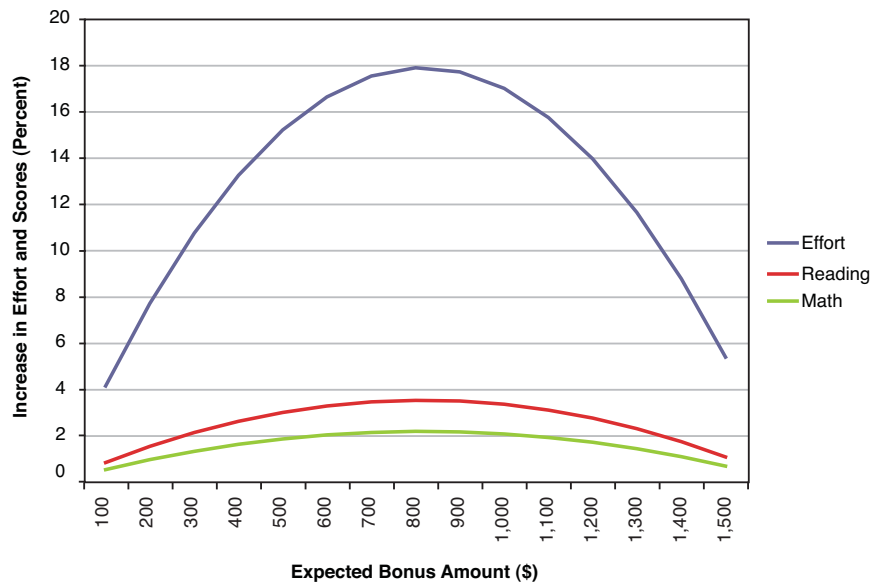
Of course, increased effort is nice, but this matters only if it translates into more learning. So do students learn more with motivated teachers? Again, the answer is yes. A highly motivated teacher will raise her students' standardized test scores by a significant amount. An average teacher who is sufficiently motivated in the North Carolina incentive system raises her students' average reading scores by more than 3.5 percent of a standard deviation and math scores by about 2.2 percent of a standard deviation. For an elementary school teacher of twenty students, the bonus program spends an average of $6.25 to raise the performance of one student in one subject by 1 percent 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.

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 the teacher must surpass 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 incentive 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 every dollar we spend will have the most bang for the buck. Economic intuition tells us that we can get rid of inefficiencies from free-rider effects by evaluating bonuses at the individual level. This seems to be a powerful case 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 the state may have backed into a more effective system than individual-level incentives.

SOURCE: Authors' calculations.

As noted earlier, 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. Both groups of teachers lack initial motivation because they are too far away from the bar set by the government. The high-ability teacher is too far above the bar, allowing her to coast and 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 is not 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 tortoise-and-hare 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 to low-ability teachers, their average score declines. While previously coasting to the bonus, they will now have to pull much harder to make it over the bar, all the while dragging the extra weight of 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
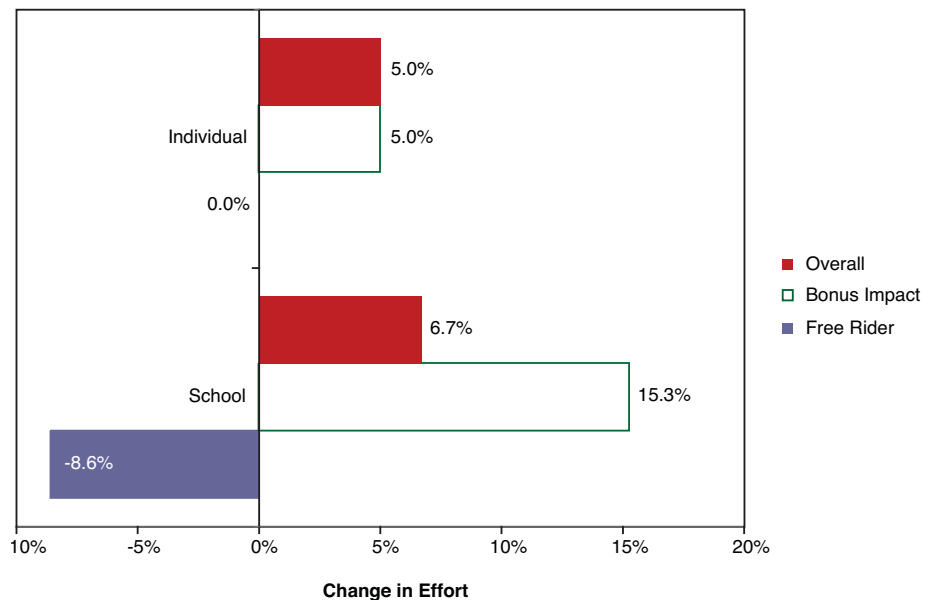
bonus, if they put in extra effort. This induces both groups of teachers to try harder.

So, which is better: school or individual incentives? In North Carolina, it appears that changing from 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, the free-rider effect is eliminated. At the same time, the change introduces the tortoise-and-hare effect by pushing most teachers away from the state standard. These two effects pull in opposite directions. Like a tug-of-war or seesaw, 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 North Carolina elementary school with about thirty-five full-time teachers. In the group-incentive regime, teachers who ignored the free-rider effect expanded their effort by about 15 percent. The free-rider effect saps more than half of this expansion, leading teachers to exert just 6.7 percent more effort on average. The individual-incentive regime eliminates the free-rider effect, but because a higher proportion of teachers views bonus receipt as either a sure thing or an unattainable goal, the average impact on effort is in fact lower, just 5 percent.

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

To be fair, there is one potential method of reducing the tortoise-and-hare effect. Rather than implementing an all-or-nothing bonus, districts could offer teachers a continually varying performance-based salary supplement. Each incremental gain in student achievement would be associated with an incremental increase in teacher pay.

FIGURE 3
SCHOOL VERSUS INDIVIDUAL INCENTIVES

SOURCE: Thomas Ahn, "The Missing Link: Estimating the Impact of Incentives on Effort and Effort on Production Using Teacher Accountability Legislation" (working paper, University of Kentucky, Lexington, KY, 2009).

The problem with such a scheme, of course, is that it magnifies the various problems associated with individual-level schemes outlined above. Continually 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 offering teachers incentives 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 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 years before rewarding deserving teachers, dismissing ineffective ones, or devoting attention to those who could 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 one. Moving to individual incentives increases each teacher's distance from the bar, introducing a tortoise-and-hare 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 do not 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.

Notes

1. See, for example, Jesse Rothstein, "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *Quarterly Journal of Economics* 125 (2010): 175–214.

2. A complete description of the bonus program and the formulas used to set each school's threshold can be found in Jacob L. Vigdor, "Teacher Salary Bonuses in North Carolina," in *Performance Incentives: Their Growing Impact on American K–12 Education*, ed. Matthew G. Springer (Washington, DC: Brookings Institution Press, 2009).

3. Ibid.