



From Data to Bonuses: A Case Study of the Issues Related to Awarding Teachers Pay on the Basis of Their Students' Progress

November 2008

In “From Data to Bonuses: A Case Study of the Issues Related to Awarding Teachers Pay on the Basis of Their Students' Progress”— a paper presented at the National Center on Performance Incentives research to policy conference in February — Daniel McCaffrey, Bing Han, and J.R. Lockwood of the RAND Corporation discuss the complex process of designing a system to award teacher bonuses on the basis of student achievement results.

Using a case study approach, the authors recount the step-by-step decisions that accompany designing such a performance pay system for teachers: (1) creating a student achievement database; (2) choosing measures of teacher performance; and (3) establishing performance thresholds for awarding bonuses. The authors conclude that the process is challenging, and each decision may have profound implications for the motivating effect of the performance pay program.

Creating a Student Achievement Database

In order to examine strategies for creating a student achievement database, the authors conducted a “hypothetical” case study, using real administrative data related to 340 middle-school mathematics teachers over a two-year period. They documented

the decision making processes involved in choosing the teachers, courses, and students to be included in the database.

Determining Which Teachers to Include

One of the first tasks in designing an incentive system is determining which teachers are eligible for bonuses. Assuming administrative data will be used, bonuses will likely be restricted to teachers of tested subjects and grades at the time of evaluation. For each teacher, a decision must be made about which assigned subjects and grade levels constitute substantial enough a portion of his/her job responsibility that results of standardized student assessments can accurately be attributed to his/her instruction.

Additionally, because objective standards for teacher performance generally do not exist, most performance measures and compensation decisions depend on a teacher's performance relative to some set of peers who are also accountable for their performance. Hence, a decision to include a teacher or group of teachers (e.g., special education teachers, reading specialists) establishes a particular reference for measuring the performance of all teachers. For example, including measures of student performance for teachers who only instruct students receiving

special services may create a distorted relative standard of teacher performance.

Identifying Relevant Courses

Next, the subject courses to be included in a performance pay system must be identified. The authors explain that it was relatively simple to determine the range of mathematics courses to include in evaluations of student achievement; however, they suggest that other subjects are less well defined and may pose a greater challenge. For example, which courses would be included when evaluating English or language arts? Would a speech class be included? What about courses such as social studies and science?

Another challenge faced by the researchers was that some students' test results in a given subject during a school year were linked to multiple teachers. In the authors' case study, they could not determine if enrollment of a student in multiple courses at the same school implied the student was enrolled in the courses sequentially or simultaneously. This uncertainty made assigning the proportion of instruction provided by each teacher difficult and required more detailed data to support estimates of teacher performance.

Selecting Student Data for Inclusion

The authors also had to establish parameters for deciding which students would be included in the database. In particular, a decision was made as to the number of days a student must be in a teacher's classroom for his or her learning to be reasonably attributable to the teacher. This decision poses several policy challenges. First, identifying students who do not "count" toward performance might result in negative incentives where teachers focus attention away from these students toward other students. Alternatively, including students with very little time in the classroom might undermine the credibility of the performance measurement system, as teachers may feel that they have not had adequate opportunity to impact those students' results. Establishing these parameters must balance such competing demands.

Choosing Measures of Teacher Performance

Specific measures of teacher performance must also be determined for the design of a performance pay system. The authors compared and evaluated various statistical models for estimating teacher performance. Despite their thorough analysis of many statistical measurement methods, they found each of the estimates to be potentially biased. Moreover, they could not rule out the possibility that every performance measure was incomplete. Given this reality, the authors had to weigh the advantages and disadvantages of each method before choosing which would be the most appropriate measure of teacher performance for their program design. They recognize that, while any measure may have its imperfections, it is possible to identify which method has the most promise to reflect program objectives. Additionally, they emphasize that the choice of measurement method has implications for the distribution of bonuses and associated motivational effects on teachers.

A detailed discussion of each model, along with its promises and limitations, is provided in the paper. For example, some performance measures had the unintended consequence of reducing the expected bonuses for the most effective teachers and increasing the expected bonuses for the truly least effective teachers. Other measures resulted in variable bonus amounts for teachers of comparable quality teaching in schools or classrooms with different student populations. Still other methods led to small percentages of teachers receiving a bonus. The authors note that not only do different levels of teacher quality affect estimates of teacher performance, but the past performance of a teacher's students also impacts these estimates. For example, some performance measures may result in overestimates of teacher performance for those instructing students with high prior test scores.

Establishing Performance Thresholds for Awarding Bonuses

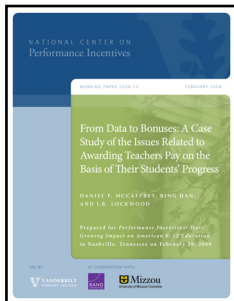
Regardless of the statistical method chosen for measuring teacher performance, the authors note

that decisions must also be made about the most appropriate performance criteria or thresholds for awarding bonuses. For example, in some performance pay designs, performance expectations are established and any teacher exceeding that threshold receives a bonus. Other award distribution methods might add a statistical significance test, so that bonuses are only awarded to those who are significantly over a given threshold. These decisions must be made carefully as they can contribute to award distribution distortions.

Each criteria choice implies a cut-off point, with some borderline teachers receiving a bonus while others do not. These methods lead to several questions with which policymakers and program designers must grapple, including, How much variance in statistical accuracy is acceptable, particularly at those points close to the cut-off? Is it better to set lower thresholds that might mistakenly reward low-performing teachers, thereby increasing the cost of the program? Or is it more detrimental to the success of the program if the system mistakenly fails to recognize and reward high-performing teachers?

Conclusion

According to McCaffrey et al., creating a performance pay system using administrative data is a challenging task. Employing a case study approach, they document their experience simulating the design of a performance award system for middle-school mathematics teachers. In the process, the authors directly address issues endemic to the use of student test results and class assignment data for the determination of teacher performance bonuses. The authors note that the many challenges encountered during this design process are not new to researchers, as they often accept a certain amount of error in their statistical analyses. However, such errors in the context of making high-stakes decisions about teacher pay must be handled with the utmost care and discretion. Therefore, the authors suggest that future research must provide information about how teachers will respond to these variations in performance pay designs to determine the best strategies for efficient and effective incentive systems.



This research brief describes work published by the National Center on Performance Incentives in “From Data to Bonuses: A Case Study of the Issues Related to Awarding Teachers Pay on the Basis of Their Students’ Progress,” by Daniel McCaffrey, Bing Han, and J.R. Lockwood, Working Paper 2008-14. The National Center on Performance Incentives is a research and development center funded in part by the United States Department of Education’s Institute of Education Sciences (R305A06034). The views expressed in this research brief do not necessarily reflect those of the sponsoring agencies.

The National Center on Performance Incentives is led by Peabody College of Vanderbilt University in partnership with the RAND Corporation and the University of Missouri-Columbia.