

Abstract Title Page
Not included in page count.

Title: Team Pay for Performance: Experimental Evidence from Round Rock's Project on Incentives in Teaching

Author(s): Daniel F. McCaffrey, John F. Pane, Matthew G. Springer, Susan F. Burns, Ann Haas

Abstract Body

Limit 5 pages single spaced.

Background / Context:

Description of prior research and its intellectual context.

What is the impact of performance pay on teacher practice and student achievement? Are individual or group-based awards the most effective means of rewarding teachers? Does performance-related pay influence teacher retention and recruitment, and if so, do effective teachers in one setting remain so in another to which they are recruited? These are timely and highly important questions for policymakers and practitioners who must consider how to best use limited resources in tight economic times with ever-increasing expectations for student achievement gains.

As is well documented, teachers are one of the most important contributors to student learning. Yet school systems spend over half their budgets paying teachers based on factors (i.e., years of experience, level of education) that are minimally related to their ability to raise students' academic performance. Nonetheless, numerous questions remain about the best alternative to the traditional single salary schedule. The research and policy communities continue debates about the promising practices and pitfalls for design and implementation of educator compensation reform, such as pay for performance, but these lessons have not always been grounded in defensible evidence.

One part of the debate around performance pay is the relative merits of individual versus group rewards. Individual awards are often valued because individual efforts are directly evaluated and rewarded, removing the possibility of "free riders" or concerns about free riders damping the effects of the compensation intervention. Alternatively, rewarding groups such as schools or teams of teachers within schools is valued because it supports collegiality among teachers, which is often cited as essential for effective schools, and because it may actually foster greater motivation as teams members push each other or the desire to support the team leads teachers to make greater changes to their teaching and to improve student learning. However, there is no empirical evidence to support hypotheses about the possible positive or negative effects of individual or group rewards.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

This paper presents the results of a rigorous experiment examining the impact of pay for performance on student achievement and instructional practice. This study, conducted by the National Center on Performance Incentives, examines a pay-for-performance program in Round Rock (Texas) which distributed performance awards to teachers based on a team's contribution to student test score gains.

The research questions are:

1. Does the opportunity to earn bonus on the basis of student achievement of taught core subjects of English language arts/reading, mathematics, science, and social studies by the

teacher and his or her teammates effect teachers' attitudes about compensation and teaching, and affect their teaching practices

2. Does the opportunity to earn bonus on the basis of student achievement of taught core subjects of English language arts/reading, mathematics, science, and social studies by the teacher and his or her teammates affect student achievement

Setting:

Description of the research location.

The study was conducted in the nine middle schools in Round Rock Independent School District (ISD) that were in operation in the 2008-09 and 2009-10 school years (a new middle school was opened for the 2010-11 school year). Round Rock ISD, a TEA Recognized School District, is located in southern Williamson County and northwest Travis County and includes the City of Round Rock and portions of the City of Austin and the City of Cedar Park. The district includes high tech manufacturing and urban retail centers, suburban neighborhoods, and farm and ranch land. The district enrollment in 2009-2010, was roughly 43,000 students attending the district's four high schools, nine middle schools, 30 elementary schools, one ninth grade center and two alternative learning centers. During the past five years, the number of students has increased by nearly 14%, and enrollment continues to grow by more than 1,200 students per year. The student population is approximately 8.7% African American, 10.7% Asian, 30% Hispanic, 46.2% White and 0.6% other.

Population / Participants / Subjects:

Description of the participants in the study: who, how many, key features or characteristics.

The study includes teachers on 78 teams of teachers teaching core subjects to students in grades 6 to 8 in nine middle schools. These are all the teams in the schools. Teams include language arts/reading, mathematics, science, and social studies teachers. Some teams also include special education teachers and specialists for English language learners. In the 2008-09 (year 1) school year, there were 371 teachers on the teams in the study. Not all teachers are part of team but most of the core subject instruction is provided by the team teachers.

Some teachers teach off team for a small proportion of their students. For instance mathematics teachers may teach a section with students from two teams or teach a section of students from another team. Off team teaching is rare in subjects other than mathematics.

There are roughly 8,440 students taught by these teachers in the year 1. Of these 8,361 were determined to be part of the teachers' team as defined as being taught two or more core subjects by the team's teachers. Our results focus only on students on a teacher's team.

Intervention / Program / Practice:

Description of the intervention, program or practice, including details of administration and duration.

Teachers in the intervention group were notified early in the school year that their team would be rated on the progress of its student in the four core subjects and the top 25% of teams in each

grade would receive a bonus in the range of \$4,500 to \$7,500 depending on the number of teachers on the winning teams.

In the summer following the school year, students were linked to teachers and teams via district administrative data. Value-added models (linear regression models with fixed team effects and student controls including prior test scores) were then used to estimate a performance measure for each team on each subject. All test scores used in the value added models were standardized to normal curve equivalents internal to the district. This placed scores from all subjects on the same score range. This allowed for creating an overall team performance measure equal to the average of the individual core subject measures.

State accountability tests were used for measuring progress in mathematics and reading for all grades and for social and social studies for grade 8 students. District administered “benchmark” assessments given in the last months of the year were used to measure grade 6 and 7 student progress in science and social studies.

Teachers received detailed reports with performance measures by subject and overall for the team as well as ranks for the team. The reports compared the performance of the team to other teams so teachers had a reference distribution for judging their performance and the average improvement in their students’ scores that would have been required to have won a bonus.

Research Design:

Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

Teams were randomized to either the bonus intervention or control condition using a block randomized design. Blocks were defined by grades within school. Within each block there were multiple teams. When there was an even number of teams half the teams in each block were randomized to treatment and half to control. In blocks with 3 teams (no blocks had more than four teams) two teams were randomly assigned to treatment or control and the remaining team was assigned to the other condition. The number of treatment and control teams was fixed at each grade level and randomizations were designed to balance the number of treatment teams across blocks to ensure the balance was met.

The district and schools provided rosters of teachers on each team and teachers were notified of their team’s assignment. Students were assigned to classes by the schools and student outcomes are analyzed according to the team’s assignment. Because assignment to experimental conditions occurred after the start of school nearly all students had their class assignments prior to teams knowing their experimental condition.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.

This paper focuses only on the year 1 effects estimated with the year 1 data.

The outcomes of interest are teacher attitudes and practices and student test scores. Teacher variables were collected via surveys during spring of year 1 and fall and spring of year 2. All teachers (treatment and control) received a \$300 stipend for completing the surveys.

The effects of the intervention on student achievement in reading and mathematics were measured by the Texas Assessment of Knowledge and Skills (the high-stakes accountability test) and the Stanford Achievement Test Series, Tenth Edition (Stanford 10), which was administered by the school specifically for this project. Effects of the intervention on student achievement in science and social studies were measured by the Stanford 10.

We fit a hierarchical linear model to estimate and test the intervention effect on student achievement. Level one models individual student outcomes as a function of a team component, and pre-intervention student variables including prior achievement scores. Level two models the team component as a function of an indicator for the intervention group, the randomization block, aggregate student pre-intervention characteristics, and random component for team.

The primary model includes a single overall intervention effect for all three grades. In secondary models we include separate effects by grade.

We used model based Wald tests to test the null hypothesis of no treatment effects. We also use randomization tests (Efron and Tibshirani, 1993) to test the null hypothesis and verify the parametric assumptions of the model.

We conducted separate analyses for each student achievement measure.

We also used a hierarchical linear model to estimate and test intervention effects on teacher attitudes and practices. Level one models teacher responses as a function of subject taught and a team component. Level two models the team components as a function of an indicator for the intervention group, the randomization block, and a random component for team. We used both model based and randomization tests of the null hypothesis of no intervention effect.

Findings / Results:

Description of the main findings with specific details.

Preliminary results find no overall intervention effect on any of the student achievement outcomes. The effects sizes are typically very small with small standard errors. At the grade level there are some significant effects. However, these are sensitive to the outcome and are currently under ongoing investigation.

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

The findings of this experiment corroborate those of the recent POINT experiment in Nashville. Offering teachers a large monetary reward on the basis of their students' progress for a small number of years of eligibility and without any addition supports does not result in teacher

improving their students' progress. Offering the bonus to teachers on teams on the basis of team's students' progress did not lead to different results than offering the bonus to teachers for the progress of their own students.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Efron, B. & Tibshirani, R. (1993). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.

Appendix B. Tables and Figures
Not included in page count.