Abstract Title Page
Not included in page count.

Title: Teaching to the Tails: Teacher Performance Pay and the Distribution of Student Achievement

Authors and Affiliations:
Prashant Loyalka, Stanford University, loyalka@stanford.edu
Sean Sylvia, Renmin University of China, sean.sylvia@gmail.com;
Chengfang Liu, Chinese Academy of Sciences, cfliu@igsnrr.ac.cn;
James Chu, Stanford University, jchu1225@stanford.edu;
Scott Rozelle, Stanford University, rozelle@stanford.edu

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

Growing evidence suggests that teachers in developing countries often have weak or misaligned incentives for improving student outcomes. In response, policymakers and researchers have proposed performance pay as a way to improve student outcomes by tying concrete measures like achievement scores to teacher pay (Lazear, 2003).

While evidence from randomized experiments generally indicates that performance pay programs are effective at improving student achievement in developing countries, there has been considerable variation in how much these programs affect student achievement. For example, a performance pay program for teachers in India had positive/long-lasting impacts on achievement scores (Muralidharan and Sundararaman, 2011; Muralidharan, 2011). Earlier studies from other developing countries also showed positive but much smaller impacts (Glewwe et al., 2003; Lavy 2009).[†]

One potential reason there may be considerable variation in impacts is because each performance pay program was designed differently. Although, taken together, the evidence suggests that performance pay holds promise for improving student achievement in developing countries, the variation in impact suggests that there are more or less effective ways to design performance pay. Because existing studies were conducted in various contexts and differ in a number of dimensions, at present we are not able to determine how specific design features affect the results. The first question pursued by our study is therefore the following: How should a school system design performance pay programs to maximize gains in student achievement?

Another largely unanswered question is how performance pay programs affect the achievement of different types of students. There are strong theoretical reasons to believe that certain types of teacher incentives could benefit students at certain parts of the achievement distribution (e.g. lower achieving students) more than other students (Neal, 2011). However, few if any studies explore how different performance pay programs affect students at certain segments of the achievement distribution more than others.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

We focus on a major design feature of performance pay programs that we believe has never been evaluated systematically in an experimental context and may have large implications. The design feature is alternative ways of linking performance pay to student achievement. For example, in many developing countries, such as China, teachers have traditionally been rewarded for the *levels* of their students' achievement (e.g., the high the average achievement scores, the more a teacher was paid). However, a trend in performance pay programs in developed countries is to reward teachers for average *gains* in achievement scores. Performance pay for *gains* could improve student achievement more than performance pay for *levels*, since gains better reflect

[†] In contrast, experimental studies of performance pay for teachers in developed countries show little or no impact on student achievement (Fryer, 2011, Springer et al., 2010).

teacher effort (Hanushek et al., 2010). We, therefore, first consider examine whether linking teacher pay to student achievement *gains* is more effective than linking pay to *levels*.

Beyond using student achievement gains or levels, another way of linking performance pay to student achievement *adjusts for student background when creating performance measures*. Specifically, researchers have suggested using performance pay designs that relate teacher rewards to the achievement gains of their students within appropriately defined comparison sets (Neal, 2011). In particular, a *pay for percentile* program has been shown (theoretically) to discourage teachers from focusing on children at the higher and middle segments of the achievement distribution at the expense of children at the lower end of the distribution (Neal, 2011). While the theory behind *pay for percentile* is compelling, it remains empirically untested.

With these alternative ways of linking performance pay to student achievement in mind, the goals of our study are to: (a) examine the impacts of different teacher performance pay designs on student achievement, both for the average student and for students across the baseline achievement distribution; (b) examine the mechanisms through which different teacher performance pay designs affect student achievement (for the average student and for students across the baseline achievement distribution).

Setting:
Description of the research location.

We conducted a large-scale randomized experiment in rural, northwest China.

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

We randomly sampled 216 schools from 16 nationally-designated "poverty" counties in Yulin Prefecture (Shaanxi Province) and Tianshi Prefecture (Gansu Province). In each school, we randomly sampled one grade 5 mathematics class (that by the time of the intervention was a grade 6 mathematics class). Altogether, we sampled 8,892 students and their grade 6 mathematics teachers.

Intervention / Program / Practice:
Description of the intervention, program, or practice, including details of administration and duration.

Schools (grade 6 classes in schools) were exposed to four different treatments: (a) control (no teacher incentive pay); (b) teacher incentive pay based on ***levels*** (Grade 6 teachers received a performance pay contract stipulating rewards based on student achievement levels on endline tests); (c) teacher incentive pay based on ***gains*** (Grade 6 teachers received a performance pay contract based on student achievement gains from baseline and endline tests); (d) teacher incentive pay based on ***pay-for-percentile*** (Grade 6 teachers received a performance pay contract stipulating rewards based on *student growth percentiles*).

Performance pay contracts (the treatments described above) were given to teachers in September 2013. Teachers (depending on which treatment arm they were assigned to) received a detailed training on their performance pay contract.

Research Design:
Description of the research design.

To test the impacts of the different teacher performance pay designs discussed above, we designed a cluster-randomized controlled trial. In this trial, schools were randomly allocated to 4 different treatment arms. The size of each treatment arm is as follows:

A. Control	A. 52 schools
B. Levels incentive	B. 54 schools
C. Gains incentive	C. 56 schools
D. Pay for percentile incentive	D. 54 schools

Note that the number of schools differ per treatment arm because our randomization was stratified by counties that had varying numbers of schools.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

Prior to the beginning of the trial (which took place when sample students were in their sixth grade year), two baseline survey waves collected information on students. The first wave of the student baseline survey was conducted in October 2011 (when sample students were at the start of their 5^{th} grade year) and the second wave was conducted in May 2012 (at the end of sample students' 5^{th} grade year). During the each wave, we collected detailed information about student and household characteristics (such as age, gender, parental education, parental occupation, family assets, and siblings). Students were also given a 30 minute standardized exam in math in each wave. During each survey wave, we also collected school-level information from school administrators. For example, we collected information on school enrollments, facilities, and distance from the county seat.

A baseline survey of teachers was also conducted in September 2013 (at the start of grade 6 for the students in our analytical sample). The survey collected information on teacher background, including information on teacher gender, ethnicity, age, teaching experience, teaching credentials, attitudes toward performance pay, and current performance pay. The teacher survey also included psychometric scales to measure social preferences including prosociality and inequality aversion. We also asked the teacher to indicate which of the sixth grade students he or she was teaching and subjective expectations about each student's potential achievement gains. The teacher baseline survey took place before we provided the grade 6 teachers with performance pay contracts (in October 2013). Note that control group teachers did not receive a performance pay contract.

We conducted our endline (post-treatment evaluation) survey in May 2014 (at the end of grade 6 for the students in our analytical sample). The endline survey collected detailed information from students, teachers, and school administrators. The information collected from students included student, teacher and household behavioral responses to the teacher incentive pay program (e.g. perceptions of teacher care/effort, teachers ability to manage classroom, teacher practices (towards student), attitudes about math (math anxiety, math self-concept), time spent on math studies each week, class ranking in math, curricula exposure in math, work in other classes outside of math, parent involvement in schoolwork, seat position in class, etc. The information collected from teachers was similar to the information collected during the baseline

survey. A principal survey was also used to collect additional information on performance pay policies and attitudes toward performance pay.

Findings / Results:
Description of the main findings with specific details.

Main impacts on achievement: Only "pay-for-percentile" incentives had a positive, statistically significant impact on average student achievement. Teacher incentives based on "levels" or "gains" were ineffective.

Secondary impacts on behavior: Pay-for-percentile incentives appeared to be effective because they changed behavior: (a) Teachers exerted more effort in the classroom, covered more challenging material and engaged students to improve their math performance; (b) Teachers were more likely to be in contact with parents about the progress of their children. (c) Students were more likely to see the importance of mathematics for their future studies and careers. By contrast, "levels" incentives appeared to be ineffective because they did not lead teachers to make changes in their behavior. Teachers that received "gains" incentives changed their behavior more than teachers with "levels" incentives but less than teachers with "pay-for-percentile" incentives.

Heterogeneous impacts of different incentive designs on students across the achievement distribution: "Gains" incentives led teachers to only focus on certain types of students which led to negligible learning (on average) across all students. By contrast, pay-for-percentile incentives led to improvements for low, medium and high performers alike. As a result, pay-for-percentile incentives led to score gains across all students (on average).

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

Our results show that the design of incentive pay matters. The results may have important implications for how Teacher Performance Pay Policy can be implemented in China and in other developing countries.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Anderson, M.L., 2008. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. Journal of the American Statistical Association 103, 1481–1495.

Betebenner, D. W. (2011a). A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth trajectories/projections. The National Center for the Improvement of Educational Assessment.

Fryer, R. G. (2011). Teacher incentives and student achievement: Evidence from new york city public schools. National Bureau of Economic Research.

Glewwe, P., Ilias, N., & Kremer, M. (2003). Teacher incentives. National Bureau of Economic Research.

Hanushek, E. A., Rivkin, S. G., Figlio, D., & Jacob, B. (2010). Generalizations about using value-added measures of teacher quality. American Economic Review, 100(2), 267–271.

Lavy, Victor. "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement." Journal of Political Economy 110, no. 6 (December 2002): 1286–1317.

Lavy, V. (2009). Performance Pay and Teachers' Effort, Productivity, and Grading Ethics. American Economic Review, 99(5), 1979–2011. doi:10.1257/aer.99.5.1979

Lazear, E. P. (2003). Teacher incentives. Swedish Economic Policy Review, 10(2), 179–214.

Lee, D.S., 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. Review of Economic Studies 76, 1071–1102.

Muralidharan, K. (2011). Long-Term Effects of Teacher Performance Pay: Experimental Evidence from India. Mimeo.

Muralidharan, Karthik, & Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. The Journal of Political Economy, 119(1), 39–77.

Murnane, R. J., & Cohen, D. K. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. Harvard Educational Review, 56(1), 1–18.

Neal, D. (2008). Designing incentive systems for schools. Performance incentives: their growing impact on American K, 12.

Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. The Review of Economics and Statistics, 92(2), 263–283.

Neal, Derek. (2011). PAY FOR PERCENTILE.pdf. NBER Working Paper 17194.

Oxman, A. D., & Fretheim, A. (2008). An overview of research on the effects of results-based financing. Nasjonalt kunnskapssenter for helsetjenesten (Norwegian Knowledge Centre for the Health Services).

Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. Journal of Policy Analysis and Management, 26(4), 909–950. doi:10.1002/pam.20292

Springer, Matthew G., Laura Hamilton, Daniel McCaffrey, Dale Ballou, Vi-Nhuan Le, Matthew Pepper, J.R. Lockwood, and Brian Stecher. TEACHER PAY FOR PERFORMANCE: Experimental Evidence from the Project on Incentives in Teaching. RAND, September 21, 2010.

Spybrook, J., S.W. Raudenbush, X. Liu, R. Congden and. A. Martinez (2009), Optimal Design for Longitudinal and Multilevel Research v1.76 [Computer Software].

State Council (2009). Guidelines on the Implementation of Performance Pay in Compulsory Education, January 1st, 2009.

Appendix B. Tables and Figures

Not included in page count.