You Get What You Pay For:

Why We Need to Invest in Strategic Compensation Reform

Matthew G. Springer[1]

University of North Carolina at Chapel Hill

August 2019

[1] Matthew G. Springer is the Robena and Walter E. Hussman, Jr. Distinguished Professor of Education Reform at the University of North Carolina at Chapel Hill. I greatly appreciate helpful comments and suggestions from Eric Houck, Matt Kraft, Jim Guthrie, Lam Pham, and Luis Rodriguez as well as course participants in Carolina's Accountability and Incentives Ph.D. seminar and the National Institute for Excellence in Teaching's roundtable on the future of public education. The usual disclaimers apply. Contact information: mgspringer@unc.edu.

Abstract

This article draws on recent insight regarding the distribution and mobility of highly effective teachers, student access to top-performing educators, and research on the effectiveness of strategic compensation reforms to argue that the single salary pay schedule has resulted in disturbing inequities for students and inefficiencies in resource allocation. These inequities are particularly alarming given that strategic compensation reforms hold promise for not only improving the quality of public education overall, but ensuring quality educational opportunities for students from traditionally underserved communities. Simply put, strategic compensation reform can meaningfully impact public education, and it is time this potential is recognized and utilized.

Introduction

Teacher compensation reform is an increasingly common strategy to enhance academic outcomes in the United States public elementary and secondary school system. School districts, state education agencies, and national and federal initiatives currently fund the development and implementation of programs that compensate teachers based on their performance and/or differentiate teacher pay in response to market conditions. These reform efforts are predicated on the argument that prevailing compensation practices provide weak incentives for teachers to act in the best interest of their students and that inefficiencies arise from rigidities in the single salary schedule.

The single salary schedule that dominates educator compensation today was first adopted in 1921 in Denver, Colorado, and Des Moines, Iowa. It governs pay according to two criteria thought to be most central to teacher productivity – years of experience and highest degree held. The single salary schedule leveled the playing field by paying teachers according to the same metric regardless of race, gender, or grade level and eliminating nepotistic practices disguised as merit pay. Highly predictable, the single salary schedule also eased annual salary negotiations between school boards and teacher unions, a particularly attractive outcome given the strained labor-management relations of the 1920s.

Single salary schedules contrast with pay practices in most other professions where differentiated pay or pay for performance is commonplace. In medicine, for instance, pay varies by specialty. Even within the same hospital or HMO, pay differs by specialty field (Folland, Goodman, and Stano, 2006). Similarly, in higher education there is significant variation in faculty pay by teaching field (Jaschik, 2016). Unlike the public K-12 system, collective bargaining agreements in higher education often include provisions that allow for field differentials based on external labor market conditions (Rhoades, 1998). Faculty pay structures also tend to be flexible. Starting pay is

generally market-driven and institutions jockey to match counter-offers to retain more senior faculty. Ultimately, the flexibility typically found in pay practices in other fields enables greater overall cost effectiveness.

Salary schedules would not be as wasteful if the factors rewarded – teacher experience and graduate education – were strong predictors of teacher productivity. However, surveys of the education production function literature find little support for a non-subject specific master's degree positively impacting student achievement, (e.g., Rivkin, Hanushek, and Kain, 2005; Clotfelter, Ladd, and Vigdor, 2006; Hanushek et al., 2005; Aaronson, Barrow, and Sanders, 2007). While recent research shows that teachers can and do continue to improve throughout their careers (Harris and Sass, 2011; Papay and Kraft, 2015; Ladd and Sorenson, 2015), there are other studies suggesting this improvement does not persist after a handful of years in the profession (Hanushek, 2003; Boyd et al,. 2008; Buddin and Zamarro, 2009; Winters, Dixon, and Greene, 2012).

The training, working conditions, and non-teaching opportunities for teachers differ significantly by teaching field, yet the salary schedule within a school district treats all teachers the same. On average the non-teaching opportunities for a high school teacher in a technical or scientific field are more remunerative than for elementary education teachers, yet the salary schedule dictates identical salaries, making it challenging for principals to hire teachers in certain subjects (Podgursky and Springer, 2011; Goldhaber, Krieg, Theobald, and Brown, 2016). Studies on the teacher labor market identify unsatisfactory working conditions as a primary motive for teachers to exit schools or the profession. Departure disproportionately penalizes students in schools with high concentrations of minority and low-income students (Simon and Johnson, 2015; Johnson, Kraft, and Papay, 2012; Boyd, Grossman, Lankford, Loeb, and Wyckoff, 2007; Johnson, Berg, and Donaldson, 2005).

Advocates of strategic compensation reforms like merit pay and hard-to-staff school bonuses believe they are a viable tool for motivating teachers to higher levels of performance, enticing effective educators to join or remain in the profession, and aligning teacher behaviors and interests with institutional goals. Nevertheless, a sturdy and influential base of individuals and organizations remains fundamentally opposed to modifying the single salary schedule, which has negatively colored perceptions of the prospect of strategic reform. These opponents argue there is little evidence that compensation initiatives such as pay for performance or recruitment and retention incentives make schools better. They contend that these programs can render schools less effective by crowding out intrinsic rewards and that the education system lacks appropriate measures for evaluating teacher performance.

This article inserts itself into the teacher compensation debate by drawing on recent insight regarding the distribution and mobility of highly effective teachers, student access to top-performing educators, and research on the effectiveness of strategic compensation reforms. This article asserts that while a single salary pay schedule may seem to promote equality among teachers, it results in disturbing inequities for students and inefficiencies in resource allocation. This is particularly alarming given that strategic compensation reforms hold tremendous promise for not only improving the quality of public education overall, but ensuring quality educational opportunities for students from traditionally underserved communities. Simply put, strategic compensation reform can meaningfully impact public education, and it is time this potential is recognized and utilized.

A Simple Thought Experiment: Can an Entrepreneur Successfully Pitch Public Education?

A simple thought experiment provides a useful litmus test for current compensation practices in elementary and secondary public education. Imagine for the moment that there is no such thing as public education. Also imagine that you are charged with pitching the notion of public

education to a set of investors, and the parameters you propose must reflect current policy and practice. Think of the setup akin to the hit television show *Shark Tank*. What might your pitch presentation sound like? What types of questions will the Sharks ask? Will the idea and funding arrangements intrigue the judges? Will you ultimately be successful in your pitch? Let's find out.

There is little doubt that your presentation starts off promisingly strong. You can guarantee that more than 50 million students will enroll annually. State governments compel children from between 5 and 19 years of age to enroll and attend,[2] after which they are no longer legally obligated. State governments are also going to limit the amount of competition that you face – you have a virtual monopoly. While this may be much to the chagrin of Milton Friedman and other free market advocates, it solidifies your position as the go-to resource in the K-12 education marketplace.

Enthusiasm for your pitch grows as you explain your financial model. Your revenue stream is based on an amalgam of local and state taxes along with a supplement from the federal government that accounts for around 9% of your total operating expenditures. In total, you project around $827 billion per year for current expenditures or around $12,000 on a per-pupil basis. You also explain than you plan to invest heavily in educators with approximately 80% of all current expenditures allocated to compensation (60%) and employee benefits (20%), while purchased services account for about 10% of total dollars, supplies for 8 percent, and tuition and other miscellaneous items about 2%.[3]

You also know the secret sauce of public education. Secret sauce is entrepreneurial language for the chief factor in the success of your product. Your secret sauce pitch is a powerful one. You review for the judges more than two decades of academic research that has clearly established that

[2] Compulsory school attendance laws vary by state, with age of required attendance ranging from 5 to 19.
[3] Values are based on data reported in U.S. Department of Education, National Center for Education Statistics, Digest of Education Statistics (2016).

the teacher is the single most important within-school determinant of student learning and that these effects vary considerably across teachers even within the same school. You know that a high-performing teacher can produce three times the achievement growth compared to a low-performing teacher in a single academic year (Hanushek and Rivkin, 2005). You also know that if a student from a low socioeconomic background receives successive years of highly effective instruction the student can overcome what is typically referred to as the achievement gap (Rivkin, Hanushek, and Kain, 2005). Moreover, encountering a highly effective teacher has long-term benefits, including high school graduation, college attendance, higher earnings, and better health outcomes (Chetty, Friedman, and Rockoff, 2014). Individual teachers also have a profound influence on students' socio-emotional development (Jackson, 2018; Kraft, 2019).

Your pitch of public education has clearly impressed the Sharks as the presentation shifts to a series of hard-hitting questions about your planned remuneration practices for the approximate 3.8 million public school educators you plan to hire. The Sharks ask questions like: *How do you reward and recognize high-performing teachers? Do you differentiate pay by the location of the school? Do you pay mathematics, science or special education teachers differently than other fields that have a greater supply than areas such as elementary education? What mechanisms are in place to ensure student access to highly effective teachers is not a result of the zip code or neighborhood in which they live?*

You explain to the judges that you do not plan to differentiate pay even though you know teachers are your secret sauce and their effectiveness can vary considerably within schools. You explain the need for pay equality; that you plan to compensate teachers based on a series of steps and lanes defined by years of teaching experience and the level of schooling that a teacher has completed. The judges seem a bit perplexed but you haven't lost them entirely. They ask a somewhat rhetorical question: *Clearly, years of experience and degree earned must have a strong influence on the outcomes you are trying to produce?*

You explain to the Sharks that understanding which teacher characteristics predict student learning has been elusive for you. One study shows that teacher characteristics that you can measure such as experience, education level, and credential explain only around 3% of the differences in student achievement (Goldhaber et al, 1999). As you try to clarify some of the nuance in teacher policy and teacher effectiveness literatures, a Shark interjects: *So, you are telling me that even though you know teachers are the single most important within-school determinant of student learning, and you plan to spend somewhere around 60 to 80% of your operational expenditures on teacher salaries and benefits, the way in which you propose to compensate these teachers has virtually nothing to do with one of, if not the primary outcome you are trying to produce – student learning? And, in most cases, this type of system is likely to perpetuate inequalities in access to high-quality educational opportunities?*

Waves of Compensation Reform

The Shark Tank thought experiment highlights persistent problems with the single salary teacher compensation model that has dominated the profession for nearly a century. Multiple waves of reform have attempted to strategically address this model's systematic shortcomings and attract, motivate, and reward highly effective teachers. These reforms generally fall into two buckets. The first comprises awards based upon some predetermined task or outcome, including merit pay, pay-for-performance, knowledge and skill-based pay, and career ladder programs. The second comprises labor-market-oriented awards, including hard-to-staff subject or school bonuses and recruitment and retention awards. The remainder of this section will focus on efforts since the 1980s, particularly as it relates to federal support for compensation reform.

Interest in teacher compensation reform has ebbed and flowed since the release of the landmark *A Nation at Risk* report in 1983 at a time when President Ronald Reagan asserted: "Teachers should be paid and promoted on the basis of their merit and competence. Hard earned

tax dollars should encourage the best. They have no business rewarding incompetence and mediocrity" (Johnson, 1984). Compensation reform during this era typically took the form of either career ladder programs or knowledge- and skill-based pay plans. Career ladder programs provided teachers new roles with additional pay and responsibilities, career advancement opportunities believed to encourage retention, and variation in responsibilities and activities designed to offset the monotony of curricular standardization. Knowledge- and skill-based pay programs rewarded teachers for successfully completing activities that demonstrated higher levels of expertise and understanding of exemplary practices. These activities typically included portfolio completion, dual certification, earning a graduate degree in subjects taught, or high marks on standards-based teacher evaluation.

Just prior to the turn of the century, a second wave of reform commenced with a series of pay initiatives layered on top of existing pay practices. These reforms were characterized by pay-for-performance and market-oriented strategies such as hard-to-staff school and subject bonuses, and represented a shift in design from rewarding educational inputs or differentiated roles and responsibilities toward rewarding based on student outcomes such as student test scores and/or growth on standardized assessments. This shift was accelerated by President George W. Bush's *No Child Left Behind Act* (NCLB), coupled with the poor relative performance of U.S. students on international math and science tests (Podgursky and Springer, 2006).

During this second wave of reform, performance pay was a critical element of educator compensation packages in the Denver and Houston public school systems.[4] Florida, Minnesota, and

[4] For research on Denver's ProComp initiative see Goldhaber and Walch (2012), Gonring, Teske, and Jupp (2007); Fulbeck, 2014), and Atteberry et al, 2015). For research on Houston's program see work by Imberman and Lovenheim (2015) and Brehm, Imberman, and Lovenheim (2017).

Texas allocated over $550 million to incentive pay programs that rewarded teacher performance.[5] And, from 2006-2010, Congress appropriated nearly $700 million to provide Teacher Incentive Fund grants to schools, districts, and states to develop and implement educator pay-for-performance plans (Office of the Inspector General, 2011), which ranged from large-scale reform efforts like the $59 million awarded to the REACH program in Austin, Texas (Balch and Springer, 2010) to a small scale pilot program in Milwaukee, Wisconsin, that was awarded less than $250,000.

The third wave of reform arrived in the form of the Obama administration's Race to the Top grants, which were authorized under the American Recovery and Reinvestment Act of 2009. In the summer of 2009, President Barack Obama and Secretary of Education Arne Duncan announced that $4.35 billion in competitive grants would be awarded to states and districts implementing innovative reforms such as turning around low-performing schools, making Race to the Top the largest competitive federal investment in education reform. At the heart of this call was the design, development, and implementation of educator evaluation systems that would account for student performance on standardized tests and inform personnel decisions about professional development, compensation, promotion, tenure, and certification. Similar to the second wave of reforms, these efforts centered on the measurement of outcomes, but with the added development of comprehensive teacher evaluation systems to enable rewarding educators for both outcomes and processes.

The Obama administration also doubled down on President Bush's investment in compensation reform. President Obama's 2011 budget request designated an additional $950 million for a new Teacher and Leader Innovation Fund to further support the development and implementation of performance-oriented compensation as a viable tool for motivating teachers to

[5] For research on Minnesota's Q-Comp see Sojourner, Mykerezi, and West (2014) and Nadler and Wiswall (2011). For research on initiatives in Texas see Springer et al (2012), Springer et al (2009a), Springer et al (2009b), and Springer and Taylor (2016).

higher levels of performance and for aligning teacher behaviors and interests with institutional goals. While Congress only approved an additional $370 million for new awards for the remainder of President Obama's term, more than $1 billion was spent in continuation awards from 2011 to 2016. An important feature of the Obama-era Teacher Incentive Fund was that performance-based compensation was coupled with other educator supports such as peer-to-peer coaching and job-embedded professional development (U.S. Department of Education, 2012). While pairing of training and compensation resulted, in part, from Nashville's Project on Incentives in Teaching Experiment findings, which I discuss in greater detail below, the training component was loosely implemented in practice. The potential dynamic complementarity between high-quality training and incentive pay remains a largely untapped policy lever in human capital production.

The Next Wave: Honing In On What Works and Why. Despite more than three decades of sustained interest in compensation reform and continued experimentation in many districts and states, past pay reform efforts have not always been well-conceived or adequately aligned with district and labor market realities. Many districts and states have implemented pay reform for the sake of pay reform; a solution was identified and implemented irrespective of a problem being present. After all, it is an attention-grabbing, progressive agenda that helps superintendents establish a reputation as reform-minded change agents. Other systems offered performance pay to all teachers because it made the reform politically feasible when, in fact, what the district really needed was to implement a strategy to reward and recognize highly effective educators and/or target high-performers in hard to staff schools. Still other districts and states introduced reforms that were watered down to satisfy diverse stakeholder demands and, as a consequence, either allocated awards to everyone in a school irrespective of their performance or reduced the award amount to the point where it was perceived by educators as a token.

The next (and potentially pivotal) generation of compensation reform will be distinguished by *intent, operationalization,* and *validation*. Intent means compensation reform that is aligned with organizational goals and objectives and designed to attract, retain, reward, and appropriately compensate talented educators. Operationalization refers not only to the better measurement of the multidimensional nature of teaching and learning and the framing of incentive rewards, but also to the adherence to design in the implementation of incentive pay programs. Validation captures the ways researchers, in partnership with practice, study compensation reform and leverage insights from behavioral economics and psychology to design, monitor, and refine cutting-edge pay systems. It is in this model of strategic compensation reform that the next generation of pay reform truly holds promise.

The importance of intent, operationalization, and validation has become increasingly apparent following findings from several high-profile incentive pay experiments, including the Project on Incentives in Teaching in Nashville, Tennessee (Springer et al, 2012a), the Team Project on Incentives in Round Rock, Texas (Springer et al, 2012b), and the Schoolwide Performance Bonus Program in New York City (Marsh et al, 2012; Fryer; 2012; and Goodman and Turner, 2013). To varying degrees, all three studies found that incentive pay systems did not improve student test scores, nor did they change teacher instructional practices or behaviors (Yuan et al, 2013), typically referred to as the motivational hypothesis of incentive pay. Albeit null, these findings provide a valuable roadmap to understanding the structural complexities and mechanisms that inhibit the success of compensation reform.

The Project on Incentives in Teaching (POINT) illustrates this dynamic. POINT was a three-year study conducted between the 2006-07 and 2008-09 school years. Middle school mathematics teachers who volunteered for the project were randomly assigned to treatment and control groups. Treatment teachers were eligible for financial rewards up to $15,000 per year if their

students made unusually large gains on standardized tests. The experiment was intended to test the hypothesis that rewarding teachers for improved scores would cause test scores to rise. It was up to participating teachers to decide what, if anything, they would do to raise student performance: participate in more professional development, seek coaching, collaborate with other teachers, focus their instruction on tested content, and/or simply reflect on their practices. Accordingly, POINT was also a test of the broader thesis that U.S. education suffers from an absence of appropriate financial incentives and that correcting the incentive structure would, in and of itself, constitute an effective intervention that improved student outcomes.

Results did not support this hypothesis. While the general trend in middle school mathematics performance was upward over the period of the project, overall, students of teachers randomly assigned to the treatment group (eligible for bonuses) did not outperform students whose teachers were assigned to the control group (not eligible for bonuses). Incentives appeared to have a positive effect in fifth grade during the second and third years of the experiment. However, the effect did not persist after students left fifth grade. Students whose fifth grade teachers were in the treatment group performed no better by the end of sixth grade than did sixth graders whose prior fifth grade teachers were in the control group.

While the storylines for the Texas and New York experiments had their own nuances, the bottom line remained the same. The programs did not elevate student outcomes compared to the group of students in the control group, and self-reported behaviors and instructional practices by teachers did not change in response to the incentive pay program. Some might argue that these compensation reform experiments prove merit pay reforms to be ineffective. However, these experiments represented a very narrow conception of pay reform (Springer and Balch, 2010), and the interventions were not strategically aligned with their specific and respective contexts. Rather, the POINT and Round Rock experiments were implemented to test a hypothesis championed by

some in the policy community – rewarding teachers or teams of teachers based on a value-added measure of teacher effectiveness will improve student achievement. In New York, the incentive system was overly complex, bonus payouts were determined after the fact and, for the most part, bonus payments treated teachers within the same school equally irrespective of their individual performance (Marsh et al, 2011).

The education community appears satisfied with the step and lane approach of the single salary schedule, whereby 21st century educators are being compensated by 20th century remuneration practices, reproducing 19th century social and economic norms that deprive specific classes of children a quality educational opportunity. Importantly, however, research is honing in on why reforms have failed and remedies for success though the null findings from high-profile initiatives like POINT have made it an uphill battle as popular media outlets, teacher associations, and select think tanks and advocacy organizations erroneously report that compensation reform does not work. Recent research, including a meta-analysis of the merit pay literature, studies of strategic retention bonus programs, and efforts to enhance effectiveness of merit pay through the power of framing, offer a far more positive outlook for compensation reform than is acknowledged. Ultimately, there is both opportunity and need to continue to articulate, test, and refine the most promising ideas and designs that will help define this next wave of compensation reform.

Strategic Compensation Reform Holds Promise

Flawed rhetoric has cast a long shadow over compensation as a viable reform strategy for improving public education. In the first few months of 2019 alone, major news outlets such as *Forbes*, *The Washington Post*, and *Education Week* published headlines about the failure of merit pay (Greene, 2019; Strauss, 2019; Sacks, 2019). Yet, as recent research on strategic compensation reform clearly demonstrates, this rhetoric is intertwined with and constrained by an overly simplified and

anachronistic view of compensation reform. The single salary schedule, while intended to eliminate gender- and race-based discrimination in pay practices, now promotes equally nefarious inefficiencies and inequality. This reality should signal a paradigmatic shift in how educator pay policies are understood particularly as we learn more from recent research: (a) merit pay has a meaningful, positive effect on test scores; (b) retention incentives can alleviate gross inequalities in educational opportunity; and (c) a well-designed incentive pay system offers tremendous promise.

Merit pay has a meaningful, positive effect on test scores.

In 2018, I completed a meta-analytic study of the teacher merit pay literature with two colleagues, Tuan Nguyen of Kansas State University and Lam Pham of Vanderbilt University. Meta-analysis is a statistical approach that synthesizes the results reported in a body of empirical research. This approach enables us to better evaluate whether teacher merit pay has influenced student test scores by systematically examining all studies conducted over a defined period of time. When there are a large number of studies on a particular intervention, such as teacher merit pay, meta-analytic techniques can quantify the amount of variance in estimated effect sizes across studies and illuminate how contextual factors and design features of these interventions lead to differential outcomes. While there are not enough studies to prescribe how merit pay programs should be defined, it is clear the overall impact of merit pay programs on student test scores is positive.

To conduct this work, we first decided on a set of criteria that needed to be met for a study to be included in our meta-analysis. We selected a fairly straightforward set of criteria. A study must focus on teachers and students in the K-12 setting, either domestic or international. A study must evaluate the impact of a merit pay program, and not some other form of compensation reform such as an overall salary increase. A study must report the effect of a merit pay program on student test scores. And, lastly, a study must use a rigorous, quantitative research design, such as randomized control trial or quasi-experimental designs.

We then identified candidate studies using a dynamic search string that combed through 20 commonly used economic and general social science databases, including ProQuest, JSTOR, NBER and EconLit. We also searched for "gray" literature using Dissertation and Thesis Repositories in WorldCat and ProQuest, as well as conducting a general Google search for evaluation reports of well-known merit pay programs such as the National Institute for Excellence in Teaching's Teacher Advancement Program (TAP) and the federal government's Teacher Incentive Fund (TIF). We then conducted searches on works cited in each candidate study meeting our eligibility criteria as well as prior narrative reviews on the topic by Chamberlin et al (2002), Harvey-Beavis (2003), Umansky (2005), Podgursky and Springer (2007), Viscardi (2012), and Imberman and Lovenheim (2015). We identified 19,908 records for screening which were subsequently reduced to 137 studies for full review, yielding a final sample of 45 studies which were independently coded (37 of which are included in final analytic sample).

In analyzing and summarizing that coding, we found that characteristics of merit pay programs implemented in the U.S. differ in two key ways from those implemented abroad. Merit pay programs in the U.S. are implemented for a shorter amount of time, on average (3.5 vs. 5.9 years). The average size of bonus awards in the U.S. is also not as large as pay awards abroad (10.1 vs 45.5 percent of per capita income). Additionally, the design features of teacher merit pay programs vary widely: about 17 percent of studies conducted in the U.S. report effect sizes for programs designed as rank-order tournaments, meaning that teachers compete for a fixed sum of money as is the case on the PGA tour. Twenty-eight percent of the studies report on programs that reward group performance, while 38% report on programs that use multiple measures of teacher effectiveness to determine eligibility for the pay award. As one might expect, there has been considerable variation in the design of programs studied from 1989 to 2018, the time frame for when studies had to be published for consideration in our meta-analysis.

We then examined the association between teacher merit pay and student test scores for studies conducted in the U.S., studies conducted outside of the U.S., and then all studies aggregated together. This is where the results are particularly interesting given rhetoric around the failure of teacher merit pay policies. We found that, on average, the effect of teacher participation in a merit pay program is associated with a statistically significant 0.043 SD increase in student test scores among studies conducted in the U.S., which is roughly equivalent to about 3 additional weeks of learning or 9% of the Black-White test score gap. The merit pay effect is even larger for studies conducted outside of the U.S., ranging between 0.070 to 0.215 SD. While the U.S. study specific average effect size is not as large as some other popular education interventions (for example, a meta-analysis by Kraft and Blazer (2018) on teacher coaching finds a pooled effect size of 0.18 SD on achievement), the studies in our meta-analysis do not account for lack of implementation fidelity. A factor shown to result in programs and policies being less effective or producing less-predictable responses (e.g., Durlak and DuPre, 2008; Century, Rudnick, and Freeman, 2010).

Take, for example, evidence from the national impact evaluation of the federally funded Teacher Incentive Fund program that was led by Mathematica Policy Research. Hanley Chiang and colleagues (2017) reported that 42% of teachers in treatment schools in the fourth year of implementation were still unaware they were eligible to earn a performance bonus. Teachers in treatment schools also reported that the maximum bonus that they could earn was no more than 40% of the actual maximum bonus districts awarded. Clearly, these factors influence the magnitude of the treatment effect, which for this particularly evaluation effort still found that students in TIF schools scored higher in math and reading.

Another informative piece of the study with Nguyen and Pham is that we reviewed studies examining the impact of merit pay programs on teacher mobility. While there were not enough studies to conduct a comprehensive meta-analytic study, the evidence supports the argument that

incentive pay will help to retain those educators most likely rewarded under the metered activity (what is typically referred to as the selection or composition effect hypothesis of incentive pay). Some studies also find that pay incentives have the potential to increase recruitment but the research in this area is even less developed and a number of questions remain, such as: What will it take to recruit a high-performing teacher to change schools?

Together, these findings clearly demonstrate that merit pay has a meaningful, positive effect on student test scores and that there is suggestive evidence of the important role these programs can play on teacher exit decisions. Most importantly, they indicate that the debate around merit pay should focus on the specific contextual factors and program design features that lead to differential outcomes in order to identify the most worthwhile investments.

Retention incentives can fundamentally alter gross inequalities in educational opportunity

The inequitable distribution of high-quality teachers across schools helps to explain the student achievement gap reported by many urban school systems (e.g., Imazeki, 2005; Scafidi, Sjoquist, and Stinebrickner, 2007; Goldhaber, Quince, and Theobald, 2017; Feng and Sass, 2017). Public schools vary significantly in their capacity to attract and retain teachers while work by Ronfeldt and colleagues (2013) has shown the negative impact of turnover on learning. We also know that schools with higher concentrations of low-income, non-white, and/or low-performing students have a more difficult time retaining teachers as school environments are likely driving some of these attrition patterns as well as students of color more likely being zoned to schools with poor working conditions (e.g., Johnson, Kraft, and Papay, 2011; Kraft et al, 2015; Loeb, Darling-Hammond, and Luczak, 2005). More experienced teachers can also use seniority-based transfer provisions in collective bargaining agreements and transfer rights to choose where to teach (Moe, 2009; Goldhaber, Lavery, and Theobald, 2016).

This sorting of teachers across schools further strengthens racial- and poverty-related achievement gaps. Schools enrolling children from the most disadvantaged backgrounds are more likely to be staffed by teachers graduating from less competitive colleges, teachers instructing out-of-field, and novice teachers (e.g., Lankford, Loeb, and Wyckoff, 2002; Goldhaber, Quince, and Theobald, 2018; Kalogrides, Loeb, and Beteille, 2012; Steele et al, 2015; Hocuk, 2010). Teacher effectiveness research consistently finds that novice teachers (defined here as teachers in their first or second year of teaching) produce smaller achievement gains for their students than more experienced teachers. The net result is that children enrolled in schools with high concentrations of disadvantaged students have greater exposure to less qualified instructors.

This inequitable distribution of high-quality teachers among schools is arguably a profound consequence of the single salary schedule interacting with the influence of non-pecuniary school characteristics such as the school's physical condition, principal leadership, safety, and distance from home. When pay is equalized, teacher quality is dis-equalized across schools. In order to equalize teacher quality and access to a quality educational opportunity, schools need to offset differences in non-pecuniary characteristics across schools, which disproportionately influence teachers' decision-making about where to work. Programs and policy innovations that seek to balance this disparity have shown positive results: strategically-designed retention incentive programs and policies have been cost-effective in retaining highly effective teachers, increasing the supply of teachers, increasing student test scores, and improving the quality of educational opportunity.

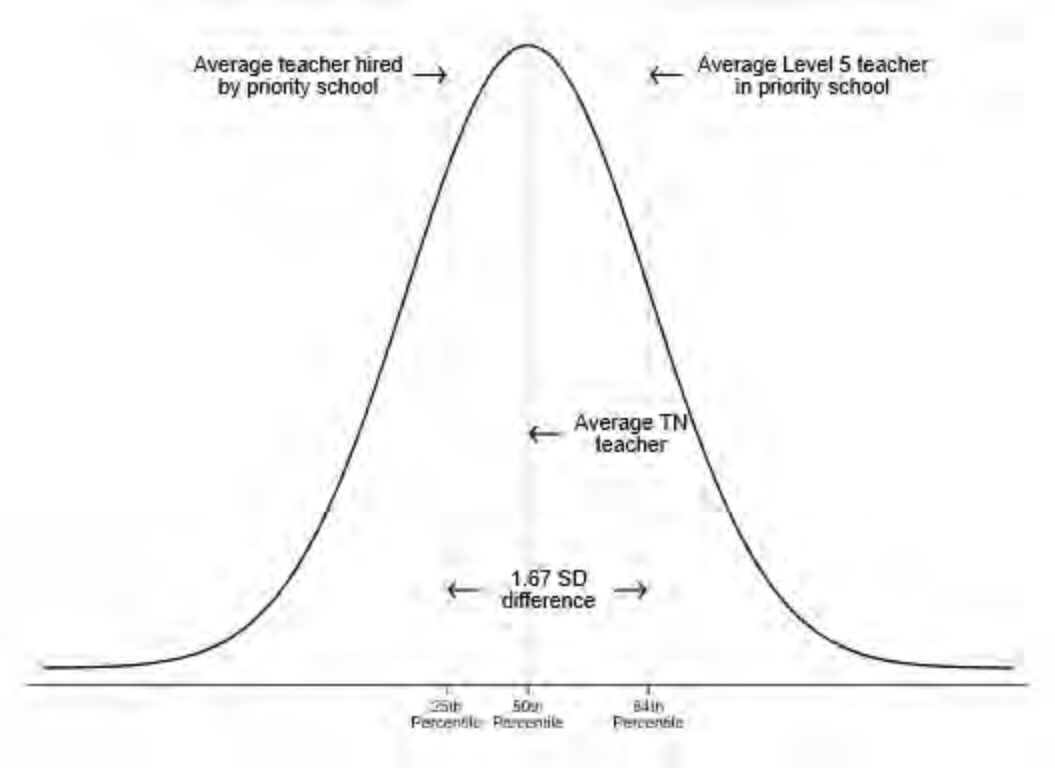
The State of Tennessee offers a prime example of how retention incentives can systematically address and disrupt gross inequalities in educational opportunity. In 2012, Tennessee identified the distribution and mobility of highly effective teachers as a public policy problem. Highly effective teachers in high-priority schools (defined here as the bottom 5 percent of schools in the state according to an accountability metric) were three times more likely to leave the school for

another teaching job than a highly effective teacher was to exit the average performing school in the state. A simple solution was proposed to combat the high rate of teacher turnover among highly effective teachers in chronically low-performing schools – offer a $5,000 retention bonus to highly effective teachers that stay in these high-priority settings.

In collaboration with Luis Rodriguez at New York University and Walker Swain at the University Georgia, we conducted an evaluation of the Tennessee Governor's Highly Effective Educator Retention Bonus Program. We leveraged a rigorous, quasi-experimental research technique called a fuzzy regression discontinuity design to estimate the causal impact of the $5,000 retention bonus on teacher retention. We found that the bonus increased the retention of high-performing educators by approximately 20%. But, this story is not just about the elevated retention rate of teachers in tested subjects and grades; it is about the differential impact of retaining the highly effective teacher relative to their likely replacement.

Picture for a moment a hypothetical distribution of all Tennessee teachers by their overall level of effectiveness, ranging from the least to most effective as you move from the left to right (see Figure 1). The average teachers in the state produce one year's worth of academic growth in their students on average, which places them at the center of the bell-shaped distribution (i.e., the 50th percentile of the distribution). The average teacher retained as part of Tennessee's highly effective teacher retention bonus is all the way to the right-hand side, at the 84th percentile of the teacher effectiveness distribution. The average teacher likely to replace this departing highly effective teacher is all the way to the left-hand side of the distribution at the 25th percentile, which equates to a difference of 1.67 standard deviations in teacher effectiveness. This is striking when you then consider that students' access to the most effective teachers can be a function of the neighborhood in which they live and the school to which they are assigned (Springer et al, 2016).

Figure 1. Teacher most likely retained as a result of the Governor's Retention Bonus Program



Adapted from Springer et al (2016).

We also considered the cost effectiveness of the selective retention bonus program for a range of plausible impacts on teachers' retention decisions, as well as the projected benefits of their retention for students in a priority school. We concluded that the cost effectiveness of the policy is primarily a function of the intervention's strength in improving the retention of a top-performing teacher in a low-performing school setting. As noted, teachers who accepted bonuses had overall teacher effectiveness ratings 1.67 standard deviations above the average replacement teacher at a priority school (Springer et al, 2016). To put this in perspective, based on estimates reported in Hanushek (2011), a single year of exposure to a teacher 1.67 standard deviations more effective can increase the total lifetime earnings of a 25-student class by greater than $835,000, net present value.

In a follow-up study, also conducted with Rodriguez and Swain, we assessed the impact of the selective retention bonus program on student achievement. We found that the retention of these

high-flier educators substantially elevated student performance in subsequent school years. Reading scores of students improved by between 0.08 and 0.12 standard deviations, or the equivalent of 40 to 57 days of additional learning assuming a traditional 180-day academic year or around one-fifth of the Black-White test score gap. The math results were similarly positive. Math scores of students improved by between 0.06 and 0.07 standard deviations, or the equivalent to 4.5 to five weeks of additional learning. However, the math score estimates were not always statistically different from zero (Swain et al, 2019).

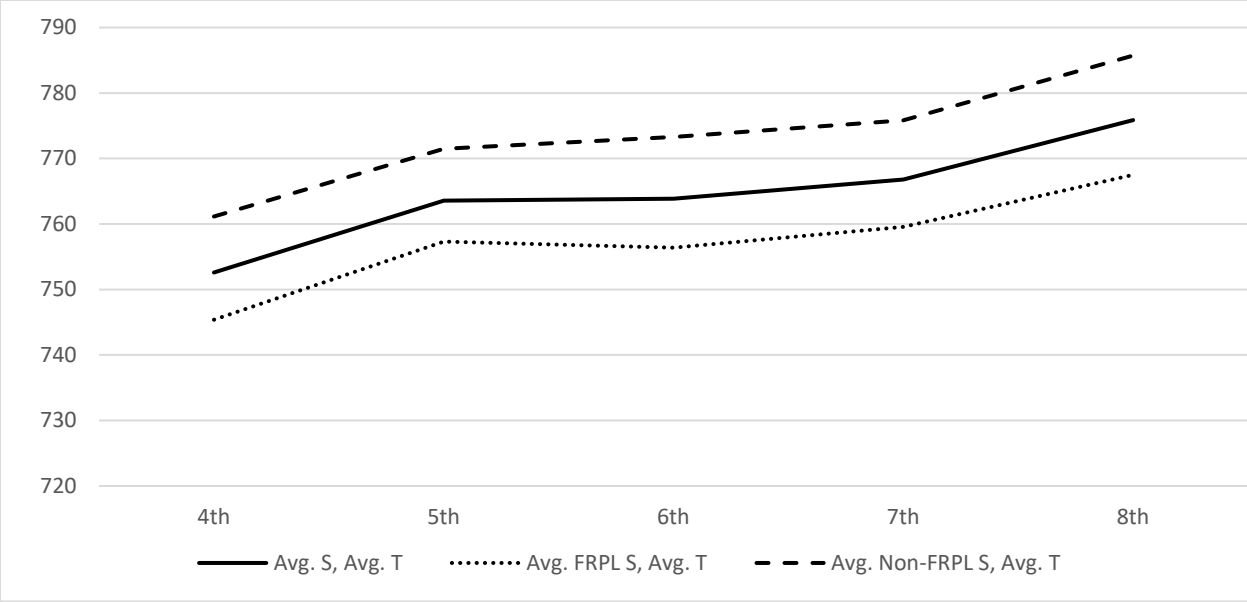
Our work in Tennessee is part of a growing number of studies to show the importance of market-oriented pay reforms as a means to equalize teacher quality across schools and provide students from traditionally underserved settings with a quality educational opportunity. In 2018, economists Li Feng at Texas State University and Tim Sass at Georgia State University published findings from an exhaustive look at Florida's Critical Teacher Shortage Program, which began in the mid-1980s and continued until terminated by the state legislature in 2011. A novel feature of Florida's program is that it offered loan forgiveness, tuition reimbursement, and (for a period of time) one-time recruitment and retention bonuses. Feng and Sass exploited changes in program coverage over time, as well as variation in the size of payments, to find that the program increased the likelihood a teacher remained in the profession and offered a cost-effective strategy to increase the supply of teachers in hard-to-staff areas. They also leveraged detailed administrative information to demonstrate how these effects varied by subject area, level of schooling, and size of bonus. All of which helps inform the next generation of pay reform.

Also in 2018, James Cowan and Dan Goldhaber at the University of Washington's Center for Education Data and Research published findings from a study of a retention bonus program in Washington that awarded financial bonuses to National Board Certified teachers in high-poverty schools. They used a regression discontinuity design to discover an elevated retention rate of

between four and eight percentage points, but no corresponding effect on student test scores. That said, the Washington and Tennessee programs differ in key ways which may explain the presence of a test score effect in one context and absence in the other. Specifically, teachers in Tennessee were awarded a retention bonus if they were rated as highly effective on the state's teacher evaluation system (one part of which is based on student test score gains) and returned to a priority school the following year. In contrast, teachers in Washington were required to complete National Board Certification, which has not been proven to be predictive of student test score gains (Chingos and Peterson, 2011; Goldhaber and Anthony, 2007; Harris and Sass, 2009).

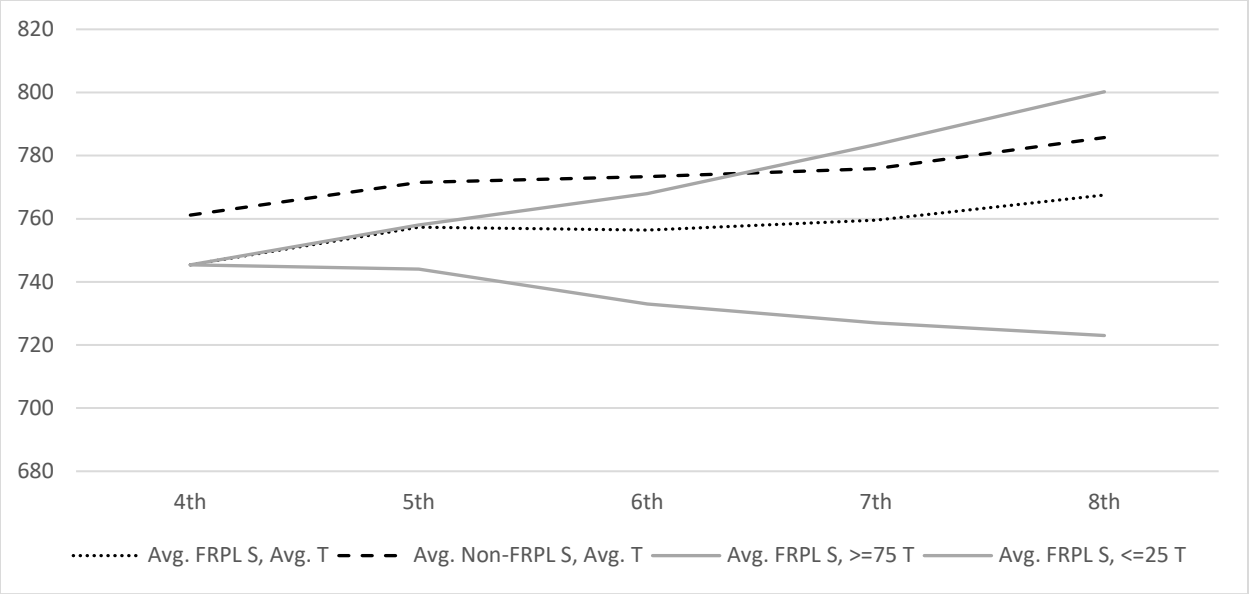
Let's now entertain a final thought experiment to illustrate: (a) the critical role that highly effective instruction can play in student learning over time and (b) the need to improve student access to quality educational opportunity. Consider, for the moment, what average math test scores look like for a cohort of students that progress from 4th to 8th grades by free-and reduced-price lunch status and their relative exposure to high- or low-performing teachers over time in a large southern state. As shown in Figure 2, the average 4th grade student taught by the average teacher in the state would score a 752 on the math assessment. As shown by the solid black line, if this average student were exposed to the average performing teacher and experienced average gains for the subsequent four school years, their math scale score would improve slightly over time, reaching a 775 by the end of 8th grade. While the levels of achievement are different, the trend is identical if we plot this trajectory for the average free-and-reduced-price-lunch student taught by the average teacher in the state (black dotted line) or for the average non-free-and-reduced-price-lunch student taught by the average teacher in the state (black dashed line). After all, we are simply assuming that a student would enroll in a classroom taught by a teacher of average effectiveness and experience the average gains of the average student in that setting from one year to the next.

Figure 2. Achievement Trajectory in Mathematics for a Cohort of Students, Grades 4 - 8



The thought experiment starts to get interesting if we focus on the test score gap between the average free-and-reduced-price-lunch student and the average non-free-and-reduced-price lunch student, which is around 16 scale score points, as shown in Figure 3. The upward sloping solid grey line represents a hypothetical scenario whereby the average free-and-reduced-price-lunch student would experience the average achievement gains for students taught by a teacher at or above the 75th percentile on the teacher effectiveness distribution from one school year to the next (that is, from the 4th to 8th grade in consecutive schools years). The student's score would increase from approximately 745 to more than 800, reaching a level higher than that of the average non-free-and-reduced-price-lunch student and resulting in an achievement gap between the two students that favors the lower-income student.

Figure 3. Achievement Trajectory in Mathematics for FRPL Students Exposed to High- and Low-Performing Teachers, Grades 4-8



What is most striking about this hypothetical, however, is the learning trajectory for the average free-and-reduced-price-lunch student if they were taught in successive years by a teacher at or below the 25th percentile of the performance distribution. As shown by the downward sloping grey line, this student's score would decrease more than 20 points, from 745 to 723. This is particularly provocative when we consider that the average teacher that would replace the departing highly effective teacher in a high-priority Tennessee school would be at the 24th percentile of the achievement distribution.

Clearly this descriptive example makes a number of assumptions, namely, it does not account for year-to-year decay in test score changes attributed to teachers, which Jacob, Lefgren, and Sims (2010) suggest are one-fifth as persistent as test score changes due to long-run knowledge. Nonetheless, given the inequitable distribution of teachers across schools, and the resultant inequitable student access to highly effective instruction, it is common sense (and urgent) that we would want to offset these market inequalities rather than perpetuate inequalities for protected classes of children.

A well-designed incentive pay system offers tremendous potential

There are a large number of complexities inherent in the design of merit pay systems: Whose performance determines bonus award eligibility? What measures will be used to monitor and appraise employee performance? Will the program reward school personnel on a relative or absolute standard? Who is part of the reformed compensation system? How and when will performance awards be distributed to school personnel? Given the complexities in the design of strategic compensation systems, and how these different design features may elicit different responses by educators, researchers have started weighing insights from behavioral economics, cognitive psychology, and prospect theory to design a new generation of pay systems that prove instructive for framing incentive pay programs to maximize outcomes of interest.

One compensation reform idea worth serious consideration is to frame incentives as losses rather than gains. In the 2010-11 school year, a team of economists from the University of Chicago conducted an incentive pay experiment in nine K-8 schools in Chicago Heights, Illinois (Fryer et al, 2018). A noteworthy feature of this project is that they not only test incentive pay in a traditional format (i.e. teachers are eligible for a bonus at the end of year based on their students' achievement), but also as an upfront payment at the beginning of the school year, part or all of which had to be returned at the end of the year depending on the percentage of students meeting the predetermined performance standard. This latter strategy is the loss aversion approach to incentive payments and is rooted in work first described by Nobel Laureate Daniel Kahneman and his collaborator, Amos Tversky in 1979.

Interestingly, Fryer and colleagues (2018) did not find an incentive pay effect when teacher performance was rewarded at the end of the school year, similar to prior experimental work in Nashville, New York City, and Round Rock. However, for the group of teachers where incentives were structured in the spirit of loss aversion, students gained between 0.20 and 0.40 standard deviations more in math, a gain that rivals the effect of widely touted and pursued reforms such as

teacher coaching and class size reduction. These effects were even larger when the incentive payments were contingent on individual teacher performance as opposed to team performance, another important feature of the experiment. While many quickly note that loss framing is unlikely to be adopted in the K-12 sector (myself being one of them), the Chicago Heights experiment was implemented in collaboration with the local teacher union and around 90% of eligible teachers volunteered to participate. It may not be as untenable as many of us have already concluded.

As an additional test of the loss aversion approach to incentive pay, Sally Sadoff, an economist at the University of California – San Diego, along with Andy Brownback at the University Arkansas, led a pair of experiments with instructors at Ivy Tech Community College, the largest postsecondary institution in Indiana. The experiment included 16 departments with over 6,000 student-course observations during the 2016-17 academic year. In the fall, instructors were allocated to either a business-as-usual control group or a treatment condition where treatment group instructors received upfront performance bonuses of $50 per student that they would have to return for each student that did not score a 70% or higher on an externally designed standardized assessment. In the spring semester, Brownback and Sadoff tested the complementarity of student incentives with the instructor incentive pay program. They cross-randomized students with existing instructor assignments which produced four treatment arms: control-group instructors, incentive-pay-only instructors; incentive-pay-only students; and the combined instructor and student incentive pay. The incentive award for students was free tuition for one summer course if they passed the spring semester end-of-course exam. Overall, they found that incentives for instructors increased exam performance by an estimated 0.16 to 0.20 SD and increased exam pass rates by 19% when compared to the control group. Students enrolled in courses taught by instructors eligible for an incentive also improved their overall course grades and increased course completion rates. Perhaps

surprisingly, Brownback and Sadoff did not find potential complementarity between the student and teacher incentives.

What is arguably most interesting about the loss aversion advancement for incentives in the education sector, as noted in Levitt and colleagues' (2016) review of the empirical literature, is that many fields have already leveraged similar insights from behavioral economics and psychology to enhance the value of their programs and policies. For example, Hossain and List (2012) found that team productivity in a high-tech manufacturing facility in China is enhanced by framing incentive payments as losses rather than gains, while Volpp et al (2008) find strong incentive effects for weight loss when participants invested their own money, which they would lose if they did not achieve individual weight loss goals. These are obviously wildly different contexts, with different underlying production functions. Regardless, the power of framing is clearly an area worthy of further consideration as we invest in and look to validate the next generation of incentive pay reforms.

Conclusion

The single salary pay schedule served an important purpose. It addressed nepotism and gender- and race-based pay discrimination that ran rampant in the 1920s. However, it also propagated gross inequity within the education system by disadvantaging traditionally marginalized communities and the students and teachers that comprise their schools. But we now have opportunity to invest in strategic compensation reforms that correct inequities in access to quality education, elevate student achievement, and improve resource efficiency without abandoning equitable opportunity for pay. It is time to make those investments.

References

Aaronson, D., Barrow, L. and Sanders, W. (2007). Teachers and Student Achievement in Chicago High Schools. *Journal of Labor Economics*, 25(1), 95–136.

Atteberry, A. and LaCour, S. (2019). Testing the Denver ProComp Theory of Action: Evidence on Intended Mechanisms for Shaping the Teacher Workforce and Student Outcomes. Working Paper. Bolulder, CO.

Ballou, D. (2001). Pay for Performance in Public and Private Schools. *Economics of Education Review*, 20(1), 51–61.

Ballou, D., and Podgursky, M.J. (1997). Teacher Pay and Teacher Quality. W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.

Brehm, M., Imberman, S. A., & Lovenheim, M. F. (2015). *Achievement Effects of Individual Performance Incentives in a Teacher Merit Pay Tournament.* National Bureau of Economic Research.

Brownback, A. and Sadoff, S. (2018). Improving College Instruction through Incentives. Working Paper. San Diego, CA.

Buddin, R. and Zamarro, G. (2009). Teacher Qualifications and Student Achievement in Urban Elementary Schools. *Journal of Urban Economics*, 66(2), 103-115.

Century, J., Rudnick, M., and Freeman, C. (2010). A Framework for Measuring Fidelity of Implementation: A Foundation for Shared Language and Accumulation of Knowledge. *American Journal of Evaluation*, 31(2), 199-218.

Chamberlin, R., Wragg, T., Haynes, G., & Wragg, C. (2002). Performance-related pay and the teaching profession: A review of the literature. *Research Papers in Education*, *17*(1), 31–49.

Chetty, R., Friedman, J.N., and Rockoff, J.E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), 2633-2679.

Chiang, H., Speroni, C., Hermann, M., Hallgren, K., Burkander, P., Wellington, A., and Warner, E. (2017). Evaluation of the Teacher Incentive Fund: Final Report on Implementation and Impacts of Pay-for-Performance Across Four Years. Washington, D.C.: Mathematica.

Chingos, M.M. and Peterson, P.E. (2011). It's Easier to Pick a Good Teacher than to Train One: Familiar and New Results on the Correlates of Teacher Effectiveness. *Economics of Education Review*, 30(3), 449-465.

Clotfelter, C., Ladd, H. and Vigdor, J. (2006). Teacher-Student Matching and the Assessment of Teacher Effectiveness. *Journal of Human Resources*, 41(4), 778–820.

Cowan, J. and Goldhaber, D. (2018). Do Bonuses Affect Teacher Staffing and Student Achievement in High-Poverty Schools? Evidence from an Incentive for National Board Certified Teachers in Washington State. *Economics of Education Review*, 65, 138-152.

Durlak, J.A. and DuPre, E.P. Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation. *American Journal of Community Psychology*, 41(3-4). 327-350.

Ehrenberg, R.G. (2004). Prospects in the Academic Labor Market for Economists. *Journal of Economic Perspectives*, 18(2), 227–238.

Feng, L. and Sass, T.R. (2017). Teacher Quality and Teacher Mobility. *Education Finance and Policy*, 12(3), 396-418.

Folland, S., Goodman, A.C. and Stano, M. (2006). The Economics of Health and Health Care, 5th Edition. Prentice-Hall, Upper Saddle River, NJ.

Fryer, R. G. (2013). *Teacher incentives and student achievement: Evidence from New York City public schools*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w16850

Fryer, R. G., Levitt, S. D., List, J., & Sadoff, S. (2018). *Enhancing the efficacy of teacher incentives through loss aversion: A field experiment*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w18237

Fulbeck, E. S. (2014). Teacher mobility and financial incentives: A descriptive analysis of Denver's ProComp. *Educational Evaluation and Policy Analysis*, *36*(1), 67–82.

Goldhaber, D., Brewer, D. and Anderson, D. (1999). A Three-way Error Components Analyses of Educational Productivity. *Education Economics*. 7(3), 199-208.

Goldhaber, D., Lavery, L., and Theobald, R. (2015). Uneven Playing Field? Assessing the Teacher Quality Gap Between Advantaged and Disadvantaged Students. *Educational Researcher*, 44(5), 293-307.

Goldhaber, D., Quince, V., and Theobald, R. (2017). Has It Always Been This Way? Tracing the Evolution of Teacher Quality Gaps in U.S. Public Schools. *American Educational Research Journal*, 55(1), 171-201.

Goldhaber, D. and Anthony, E. (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.

Goldhaber, D. & Walch, J. (2012). Strategic pay reform: A student outcomes-based evaluation of Denver's ProComp teacher pay initiative. *Economics of Education Review*, *31*(6), 1067–1083.

Golhaber, D., Krieg, J., Theobald, R., and Brown, N. (2015). Refueling the STEM and Special Education Teacher Pipeline. *Phi Delta Kappan*, December/January, 56-62.

Gonring, P., Teske, P., and Jupp, B. (2007). Pay-for-Performance Teacher Compensation: An Inside View of Denver's ProComp Plan. Cambridge, MA: Harvard University Press.

Goodman, S., & Turner, L. (2011). Does whole-school performance pay improve student learning? Evidence from the New York City schools. *Education Next*, *11*(2), 67–71.

Greene, P. (2019). Teacher Merit Pay Is a Bad Idea. *Forbes*, February 9, retrieved from https://www.forbes.com/sites/petergreene/2019/02/09/teacher-merit-pay-is-a-bad-idea/#5f3e210f4ffb.

Hanushek, E.A. (2003). The Failure of Input-based Resource Policies. *Economic Journal*, 113(485), F64–F68.

Hanushek, E. A. (2011). The economic value of higher teacher quality. Economics of Education Review, 14, 101–117.

Hanushek, E. A. and Rivkin, 2004. "How to Improve the Supply of High-Quality Teachers." Brookings Papers on Education Policy, 7-25. Brookings Institution, Washington, D.C.

Hanushek, E.A, Kain, J. O'Brien, D. and Rivkin, S. (2005). The Market for Teacher Quality. NBER Working Paper 11154. National Bureau of Economic Research, Cambridge, MA.

Harris, D.N. and Sass, T.R. (2008). The Effects of NBPTS-Certified Teachers on Student Achievement. *Journal of Policy Analysis and Management*, 28(1), 55-80.

Harris, D.N. and Sass, T.R. (2011). Teacher Training, Teacher Quality, and Student Achievement. *Journal of Public Economics*, 95(7-8). 798-812.

Harvey-Beavis, O. (2003). Performance-based rewards for teachers: A Literature Review*, UK.* Retrieved from http://www.academia.edu/download/41008714/types_of_rewards.pdf

Hossain, T. and List, J.A. (2012). The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations. *Management Science*, 58(12), 2151-2167.

Houck, E.A. (2010). Teacher Quality and School Resegregation: A Resource Allocation Case Study. *Leadership and Policy in Schools*, 9(1), 49-77.

Imazeki, J. (2005). Teacher Salaries and Teacher Attrition. *Economics of Education Review*, 24(4), 431-449.

Imberman, S. A., & Lovenheim, M. F. (2015). Incentive strength and teacher productivity: Evidence from a group-based teacher incentive pay system. *Review of Economics and Statistics*, *97*(2), 364–386.

Jackson, C.B. (2018). What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes. *Journal of Political Economy*, 126(5), 2072-2107.

Jacob, B., Lefgren, L., and Sims, D.P. (2010). The Persistence of Teacher-Induced Learning. *Journal of Human Resources*, 45(4), 915-943.

Jaschik, S. (2016). What You Teach Is What You Earn. *Inside Higher Education*, March 28, retrieved from https://www.insidehighered.com/news/2016/03/28/study-finds-continued-large-gaps-faculty-salaries-based-discipline.

Johnson, S.M. (1986). Incentives for Teachers: What Motivates, What Matters? *Educational Administration Quarterly*, 22(23), 54-79.

Johnson, S.M., Kraft, M.A., and Papay, J. (2012). How Context Matters in High-Need Schools: The Effects of Teachers' Working Conditions on Their Professional Satisfaction and Their Students' Achievement. *Teachers College Record*, 114(10), 1-39.

Johnson, S.M., Berg, J.H., and Donaldson, M.L. (2005). Who Stays in Teaching and Why: A Review of the Literature on Teacher Retention. NRTA Educator Support Network.

Kraft, M.A., (2019). Teacher Effects on Complex Cognitive Skills and Socio-Emotional Competencies. *Journal of Human Resources*, 54(1), 1-36.

Kraft, M., Blazer, D., and Hogan, D. (2018). The Effect of Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence. *Review of Educational Research*, 88(4), 547-588.

Kraft, M.A., Papay, J.P., Johnson, S.M., Charner-Laird, M., Ng, M., Reinhorn, S. (2015). Educating Amid Uncertainty: The Organizational Supports Teachers Need to Serve Students in High-Poverty, Urban Schools. *Educational Administration Quarterly*, 51(5), 753-790.

Ladd, H.F. and Sorenson, L.C. (2017). Returns to Teacher Experience: Student Achievement and Motivation in Middle School. *Education Finance and Policy*, 12(2), 241-279.

Lankford, H., Loeb, S. and Wyckoff, J. (2002). Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37–62.

Levitt, S.D., List, J.A., Neckermann, S. and Sadoff, S. (2016). The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance. *American Economic Journal: Economic Policy*, 8(4), 183-219.

Loeb, S., Darling-Hammond, L., and Luczak, J. (2005). How Teaching Conditions Predict Teacher Turnover in California Schools. *Peabody Journal of Education*, 80(3), 44-70.

Loeb, S., Kalogrides, D., and Beteille, T. (2012). Effective Schools: Teacher Hiring, Assignment, Development, and Retention. *Education Finance and Policy*, 7(3), 269-304.

Marsh, J. A., Springer, M., McCaffrey, D., Yuan, K., & Epstein, J. (2011). *A big apple for educators: New York City's experiment with schoolwide performance bonuses: final evaluation report*. Santa Monica, CA: RAND.

Nadler, C. and Wiswall, M. (2011). Risk Aversion and Support for Merit Pay: Theory and Evidence from Minnesota's Q-Comp Program. *Education Finance & Policy*, 6(1), 75-104.

Moe, T.M. (2009). Collective Bargaining and the Performance of the Public Schools. *American Journal of Political Science*, 53(1), 156–174.

Office of the Inspector General (2011). Department's Implementation of the Teacher Incentive Fund Grant Program: Final Audit Report. Washington, D.C.: Author.

Papay, J. and Kraft, M.A. (2015). Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-Term Career Improvement. *Journal of Public Economics*, 130, 105-119.

Pham, L., Nguyen, T., and Springer, M.G. (2018). Teacher Merit Pay. Working Paper. Chapel Hill, NC.

Podgursky, M. (2007). Teams versus Bureaucracies: Personnel Policy, Wage-Settings, and Teacher Quality in Traditional Public, Charter, and Private Schools. In Mark Berends, Matthew Springer, and Herbert J. Walberg (eds.), Charter School Outcomes, 61–84. Taylor & Francis, Inc., New York, NY.

Podgursky, M. and Springer, M.G. (2007). Teacher Performance Pay: A Review. *Journal of Policy Analysis and Management*, 26(4), 909–950.

Podgursky, M. and Springer, M.G. (2011). Teacher Compensation Systems in the United States K-12 Public School System. *National Tax Journal*, 64(1), 165-192.

Rhoades, G. (1998). Management Professionals: Unionized Faculty and Restructuring Academic Labor. SUNY Press, Albany, NY.

Rivkin, S.G., Hanushek, E.A. and Kain, J.F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2), 417–458.

Ronfeldt, M., Loeb. S., Wyckoff, J. (2013). How Teacher Turnover Student Achievement. *American Educational Research Journal*, 50(1), 4-36.

Rothstein, J. (2010). Teacher Quality in Education Production: Tracking, Decay, and Student Achievement. *Quarterly Journal of Economics*, 125(1), 175-214.

Sacks, A. (2019). Teachers Need More Pay, Not Merit Pay. *Education Week*, March 12, Retrieved from https://www.edweek.org/tm/articles/2019/03/12/teachers-need-more-pay-not-merit-pay.html.

Simon, N.S. and Johnson, S.M. (2015). Teacher Turnover in High-Poverty Schools: What We Know and Can Do. *Teachers College Record, 117(3).*

Scafidi, B., Sjoquist, D.L., and Stinebrickner, T.R. (2007). Race, Poverty, and Teacher Mobility. *Economics of Education Review*, 26, 145-159.

Sojourner, A. J., Mykerezi, E., & West, K. L. (2014). Teacher pay reform and productivity panel data evidence from adoptions of Q-Comp in Minnesota. *Journal of Human Resources*, *49*(4), 945–981.

Springer, M. G. (2009). *Performance incentives: Their growing impact on American K-12 education*. Brookings Institution Press.

Springer, M. G., & Balch, R. (2010). Design components of incentive pay programs in the education sector. In *Teacher Incentives and Stimuli*. Paris: Organisation for Economic Co-Operation and Development.

Springer, M. G., Ballou, D., Hamilton, L., Le, V.-N., Lockwood, J. R., McCaffrey, D. F., … Stecher, B. M. (2011). Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT). Retrieved from http://eric.ed.gov/?id=ED518378

Springer, M. G., Ballou, D., & Peng, A. X. (2014). Estimated effect of the teacher advancement program on student test score gains. *Education, Finance, and Policy*, *9*(2), 193–230.

Springer, M. G., Lewis, J. L., Ehlert, M. W., Podgursky, M. J., Crader, G. D., Taylor, L. L., … Stuit, D. A. (2010). District Awards for Teacher Excellence (DATE) Program: Final Evaluation Report. *National Center on Performance Incentives*. Retrieved from http://eric.ed.gov/?id=ED538372

Springer, M. G., Lewis, J. L., Podgursky, M. J., Ehlert, M. W., Taylor, L. L., Lopez, O. S., & Peng, A. (2009). Governor's Educator Excellence Grant (GEEG) Program: Year Three Evaluation Report. *National Center on Performance Incentives*.

Springer, M. G., Pane, J. F., Le, V.-N., McCaffrey, D. F., Burns, S. F., Hamilton, L. S., & Stecher, B. (2012). Team pay for performance: Experimental evidence from the Round Rock Pilot Project on team incentives. *Educational Evaluation and Policy Analysis*, *34*(4), 367–390.

Springer, M. G., Podgursky, M. J., Springer, C., Hamilton, L. S., Lopez, O. S., Peng, A. X., & Stecher, B. M. (2009). Texas Educator Excellence Grant (TEEG) Program. Retrieved from http://130.154.3.8/pubs/external_publications/EP20081235.html

Springer, M. G., Swain, W. A., & Rodriguez, L. A. (2015). Effective teacher retention bonuses: Evidence from Tennessee. *Educational Evaluation and Policy Analysis*, 0162373715609687.

Springer, M. G., & Taylor, L. L. (2016). Designing incentives for public school teachers: Evidence from a Texas incentive pay program. *Journal of Education Finance*, *41*(3), 344–381. https://doi.org/10.1353/jef.2016.0001

Steele, J., Pepper, M.J., Springer, M.G., and Lockwood, J.R. (2015). The Distribution and Mobility of Effective Teachers. *Economics of Education Review*, 48, 86-101.

Strauss, V. (2019). Performance Pay Drove Denver Teachers to Strike – And It Is Failing in Other Districts, Too. The Washington Post, February 12, Retrieved from

https://www.washingtonpost.com/education/2019/02/12/performance-pay-drove-denver-teachers-strike-it-is-failing-other-districts-too-north-carolina-teacher-explains-why-it-doesnt-work/?utm_term=.773e2e149628

Swain, W.A., Rodriguez, L.A., Springer, M.G. (2019). Selective Retention Bonuses for Highly Effective Teachers in High Poverty Schools: Evidence from Tennessee. *Economics of Education Review*, 68, 148-160.

Umansky, I. (2005). A Literature Review of Teacher Quality and Incentives. *Incentives to Improve Teaching*, 21.

U.S. Department of Education (2012). 2012 Teacher Incentive Fund Invites Districts to Pursue a New Vision for Human Capital Through Stronger Evaluations and Greater Professional Opportunities. Washington, D.C.: Author. Retrieved from https://www.ed.gov/news/press-releases/2012-teacher-incentive-fund-invites-districts-pursue-new-vision-human-capital-through-stronger-evaluations-and-greater-professional-opportunities

Viscardi, D. (2014). The Teacher Pay for Performance Phenomenon. *Unpublished Doctoral Dissertation. Seton Hall University.*

Volpp, K.G., Leslie, J.K., Troxel, A.B., Norton, L., Fassbender, J., and Loewenstein, G. (2008). Financial Incentive-Based Approaches for Weight Loss. *Journal of the American Medication Association*, 300(22), 2631-2637.

Winters, M. A., Dixon, B. L., & Greene, J. P. (2012). Observed characteristics and teacher quality: Impacts of sample selection on a value added model. *Economics of Education Review*, 31(1), 19-32.

Yuan, K., Le, V., McCaffrey, D.F., Marsh, J.A., Hamilton, L., Stecher, B., and Springer, M.G. (2013). Incentive Pay Programs Do Not Affect Teacher Motivation or Reported Practices: Results from Three Randomized Studies. *Educational Evaluation and Policy Analysis,* 35(1), 3-22.