The Prospects of Teacher Pay for Performance

Derek C. Briggs

Michael Turner

Charles Bibilos

Andrew Maul

University of Colorado, Boulder

May 6, 2014

CADRE Working Paper 2014-01

Abstract

The theory of action for teacher pay for performance (PFP) is based on two central premises: the motivation premise, which suggests that performance incentives motivate teacher effort, and the selection premise, which suggests that PFP systems incentivize high-performing teachers to remain in their jobs. Using data from Denver's Professional Compensation for Teachers (ProComp) system, we show that neither premise necessarily holds in practice. Under ProComp, some teachers experience declines in pay from year to year. Because teachers are strongly averse to these marginal losses in salary, they may not perceive PFP systems as incentivizing their behavior. This view is consistent with the tenets of prospect theory, and may explain why PFP systems appear to have minimal impact on student outcomes.

Introduction

Teacher pay for performance (PFP) systems are becoming increasingly prevalent in urban public school districts in the United States. Districts in New York, Houston, Chicago, and Denver have all implemented variations of PFP over the past ten years (Goldhaber & Walch, 2012). The theory of action behind this approach hinges upon two fundamental premises. The first, the "motivation" premise, is that performance incentives can motivate teachers to work harder by tying pay to student outcomes. The second, the "selection" premise, is that incentives can be established that optimize the quality of teachers who enter and remain in schools. If either or both of these premises about PFP hold, positive effects on student achievement would be an anticipated result.

In this paper we use the results from an evaluation study of one of the oldest and best-known PFP systems in the country—Denver Public Schools' Professional Compensation (ProComp) system—to provide a scenario in which neither the motivation nor the selection premises in the PFP theory of action are likely to hold. This scenario was initially motivated by focus group interviews with a total of 47 teachers who were asked to reflect upon the ProComp system's impact on teaching practices and teacher recruitment and retention. On the whole, not only were these teachers dismissive of ProComp's impact on their practices and retention, but they tended to regard the magnitude of ProComp salary incentives as "insultingly" low. This was the case even though the average teacher in ProComp earned about $20,000 in bonuses between the 2009-10 and 2011-12 school years. We argue that Denver teachers are most likely adopting a frame of reference—marginal gains in salary—consistent with prospect theory (Kahneman & Tversky, 1979; 1984), while district leadership is adopting a frame of reference—cumulative

gains in salary—more consistent with rational choice theory (Becker, 1979). This disconnect in frames of reference may explain why neither the motivation nor selection premises behind the PFP system are valid, and therefore why it comes as no surprise to find little evidence of a system-wide effect on student achievement (Goldhaber & Walch, 2012; Briggs, Diaz-Bilello, Maul, Turner, & Bibilos, 2014).

Background on Pay for Performance

Most research on teacher PFP systems cites Lazear's (2000) and (2003) studies of the Safelite Autoglass Company as an illustration of the positive expected impact of merit pay (Besharov, Podgursky, & Springer, 2007; Goldhaber & Walch, 2012; Hanushek & Raymond, 2006; Hanushek & Woessmann, 2011; Muralidharan & Sundararaman, 2009; Podgursky & Springer, 2011, 2007; Woessmann, 2011). In Lazear's initial study, the productivity of Safelite's windshield installers increased by 44% after the company switched from hourly wages to a pay-per-item system of merit pay. The new compensation system is thought to have increased productivity for two reasons. First, productivity increased because workers saw a direct relationship between their effort and their pay: when technicians completed windshield installations more quickly and efficiently, their pay increased proportionately. Secondly, since highly productive technicians earned more money than their less productive colleagues, retention rates were higher among the most productive workers. These market selection effects had as much of an impact on overall productivity as the motivational effects. The change to a performance-based compensation system was ultimately a boon to the company: the 44%

increase in output per worker was accomplished with only a 10% increase in total compensation costs per worker.

Lazear (2003) subsequently extrapolated his theories on PFP systems to educational settings, recommending that schools eliminate all policies that limit market selection effects, including teacher tenure and traditional step-based salary schedules. Further, Lazear suggested that teacher compensation based on student outcomes would replicate the motivational effect observed in Safelite's pay-per-item system, and he also argued that teachers should be paid more if they teach relatively hard-to-staff grades and subjects.

However, despite the growing popularity of PFP systems in educational settings, there seems to be little evidence that PFP has succeeded in improving student outcomes in the United States. Although studies in other parts of the world have yielded significant positive effect sizes for PFP systems under specific conditions (Muralidharan & Sundararaman, 2009; Woessmann, 2011), no recent studies of PFP systems in the U.S. have found large effects on student outcomes. By far the most authoritative study on PFP systems is the Tennessee POINT experiment, which did not yield statistically significant effects on overall student performance (Springer et al., 2010). In a quasi-experimental evaluation of Denver ProComp, Goldhaber & Walch (2012) found statistically significant effect sizes, but these effects were quite small and often negative, ranging from -0.04 to 0.05, depending on the grade level and subject tested. In addition, it was unclear how much of these effects could be attributed solely to ProComp since the authors were using a historical control group. Recent analyses of PFP systems in Chicago (Glazerman & Seifullah, 2012) and New York City (Fryer, in press; Goodman & Turner, 2011; Springer & Winters, 2009) have also failed to find statistically significant impacts on student achievement.

The results of these empirical studies raise some important questions about the validity of the motivation and selection premises, as well as their connection to student achievement. In particular, the POINT study calls into question the validity of the motivation premise. When surveyed, teachers participating in the treatment condition indicated that the large monetary incentives did not motivate them to work any harder, presumably because they had already been working as hard as they could. In contrast, the validity of the selection premise remains an open question because the anticipated impact of selection effects on student outcomes could take many years to manifest. Because most major PFP systems in public education settings have existed for a relatively short time, and because the empirical studies that have been conducted on them have either been short-term in nature (Springer et al., 2010) or were equivocal because of threats to internal validity (Fulbeck, 2013), proponents of PFP systems can still plausibly argue that market selection may ultimately have a positive impact on student performance in the long run.

Denver Public Schools ProComp System

One of the first localities to adopt a PFP system was Denver, Colorado, where the Professional Compensation System for Teachers (ProComp) was piloted in 2004 and became the primary compensation system for new teachers and those who self-selected into the system in 2006 (Gonring, Teske & Jupp, 2009). The ProComp system was among the first and most ambitious of its kind, and was funded by a large voter-approved tax levy and a U.S. Department of Education Teacher Incentive Fund grant. When ProComp was implemented, Denver Public Schools (DPS) hoped that the system, with its roughly $30 million in funds dedicated to

incentive pay, would improve outcomes for DPS students by providing extra motivation for teachers and helping the district retain its best teachers. As mentioned previously, studies so far have shown that ProComp appears to have had only a small impact on teacher retention and student outcomes (Fulbeck, 2013; Goldhaber & Walch, 2012).

The ProComp system includes 10 distinct financial incentives, divided into four broad categories: Knowledge and Skills, Student Growth, Market Incentives, and Comprehensive Professional Evaluations. The dollar amounts awarded for each of the 10 ProComp incentives are based on a pre-negotiated percentage of an overall index, which was held constant at $37,551 from the beginning of the 2009-10 academic year through the 2011-12 academic year. Awards for meeting the various ProComp incentives come in the form of either a one-time bonus or a permanent, "base-building" salary increase, depending on the incentive and the teacher's credited tenure in DPS. Of the 10 incentives currently offered by DPS, six are offered only as a one-time bonus, two are offered only as base-building salary increases, and two are applied as either a bonus or a salary increase, depending on the circumstances. The various ProComp incentives are summarized in Table 1.

Table 1. ProComp Incentive Amounts

Area of Focus	ProComp Award	Dollar Amount	
		One-Time	Base Increase
Student Growth	Top Performing School	$2,403	
	High Growth School	$2,403	
	Exceeds Expectations	$2,403	
	1 Growth Objective Met	$376	
	2 Growth Objectives Met		$376
Market Incentive	Hard to Staff	$2,403	
	Hard to Serve	$2,403	
Knowledge & Skills	Professional Development		$751*
	Tuition Reimbursement	$1000/year	
	Advanced Degrees		$3,380
Comprehensive Evaluation	Probationary Teacher		$376
	Non-Probationary Teacher		$1,127

Note. Successfully completing a Professional Development Unit only counts toward base salary during the first 14 years of credited service in DPS. After that it constitutes a one-time bonus.

Student growth incentives include four separate awards based on student achievement. Two of these incentives, Top Performing and High Growth, are awarded on the basis of school level performance. The High Growth award is based on school-wide average growth on standardized tests, as measured by Student Growth Percentiles from the Colorado Growth Model (Betebenner, 2009). The Top Performing award is based on a broad index of performance measures, including standardized test scores, attendance rates, and graduation rates. The Exceeds Expectations award is earned by individual teachers in eligible grades and subjects whose students meet Student Growth Percentile targets in math and reading; Student Growth Objective incentives are awarded to teachers who establish and achieve district-approved goals for their students. None of the other six ProComp incentives are directly tied to measures of student outcomes. Market Incentives reward teachers for working in either Hard to Staff positions, such as bilingual special education and secondary mathematics, or Hard to Serve schools, defined by DPS as schools with large percentages of students eligible for free or reduced lunch. Knowledge and Skills awards are given to teachers who complete approved professional development units

or attain an advanced degree. Comprehensive Professional Evaluation awards are awarded to teachers who successfully complete an evaluation process, generally conducted either every year or every three years, depending on a teacher's tenure with the district.

All teachers and student service professionals who are covered by the Denver Classroom Teachers Association collective bargaining agreement are eligible to join ProComp. ProComp-eligible employees include social workers, psychologists, school librarians, nurses, therapists, and intervention teachers, in addition to conventional classroom teachers. Charter school employees, however, are not eligible to join ProComp. In what follows we use the term "teachers" to refer to all ProComp-eligible educators. Our population of interest in this study consists of the 2,430 teachers who participated in ProComp and were present in all three years from 2009-10 to 2011-12.

Cumulative Salary Gains Relative to 2009-10 Base Salary

As a first step in exploring the ProComp incentive system and its impact on teachers, we computed for each teacher cumulative salary gains attributable to earning one or more of the ProComp incentives listed in Table 1. To do this, we first added the total of all one-time bonuses that each teacher earned from the 2009-10 through the 2011-12 school years, and then added earnings from base-building incentives. For example, if a teacher earned a $376 base-building professional development reward in 2009-10, the reward would be permanently added to the teacher's pay in 2010-11 and 2011-12, and would therefore increase the teacher's cumulative salary gain by $376 each year, for a total of $1,128 over the three-year period.

The average DPS teacher earned a cumulative $20,036 over and above their 2009-10 base salary, with a standard deviation of $7,228. Although a small proportion of teachers (8%) earned less than $10,000 in incentives over this time period, roughly the same proportion earned more than $30,000. From the perspective of DPS leadership, ProComp would appear to be injecting a considerable amount of money into teacher's paychecks—on average, about $6,700 per year.

Focus Groups

Variability in teachers' incentive attainment does not, by itself, guarantee that a PFP system is adequately incentivizing teacher behavior and practices in a way that will lead to improved student outcomes. Because recent studies suggest that the link between PFP incentives and student achievement is relatively weak, we conducted eight semi-structured focus groups with DPS teachers to learn more about their perceptions of the ProComp system. In particular, we wanted to ask teachers whether they felt motivated by ProComp incentives (i.e., the PFP motivation premise) and whether ProComp had an impact on their employment decisions and teaching practices (i.e., the PFP selection premise).

Our eight focus groups included a total of 47 participants in small groups of four to seven teachers. We purposely selected a diverse sample of teachers that would represent a variety of schools, subjects, and grade levels taught. Specifically, our focus groups included both veteran teachers and relative newcomers to the district, as well as a handful of teachers who had chosen not to opt in to ProComp. Out of 47 participants, 13 had non-teaching positions in DPS as counselors, librarians, therapists, intervention teachers, and nurses—all of whom were also eligible to earn ProComp incentives.

Each focus group included a protocol with standardized questions grouped by theme. However, once questions were posed, the ensuing discussions were relatively free form and participants were encouraged to openly voice their opinions. All interviews were recorded and members of the research team took notes. These notes, along with transcriptions of the recordings, were later examined by the full team to help identify commonly expressed themes. Once these themes were established, we returned to our recordings and notes and tabulated the frequency with which individual participants made statements consistent with each theme during the group interview. Hence, if it is reported that 80% of participants made a statement consistent with a particular theme, this does not necessarily indicate that the complement (100% − 80% = 20%) made statements counter to this theme. This tabulation and written reports from our field notes are the principal data sources for the results that follow.

Focus group participants overwhelmingly agreed that ProComp incentive awards are too small: 89% of participants stated that ProComp incentive sizes were small to the point of being "insulting." Nearly every ProComp incentive was criticized as "insignificant" by at least some focus group participants. In particular, teachers did not understand why the Student Growth Objective incentive ($376) was so much smaller than other incentives, and even the Hard to Serve and Hard to Staff bonuses ($2,403) were often cited as insufficient to make a difference in employment decisions.

A majority of teachers (64%) commented that ProComp has had minimal influence on their teaching practices. In one focus group, a teacher said that "we didn't go into this job for the money," and teachers expressed similar sentiments in each of the other seven groups. Many participants asserted that their students are their primary source of motivation, and that ProComp incentives do not substantially alter their work habits: "it would never be in my mind to say, if I

just work a little bit harder I'm going to get $2,300." Some teachers expressed appreciation for the extra pay earned through ProComp incentives, but agreed that their motivations are primarily linked to their students. Similarly, 77% of focus group participants stated that ProComp has had no impact on their decision to come to DPS or to remain at DPS. A number of teachers said that they chose to remain at a particular school primarily because of the sense of community that they feel in the workplace, not because of their pay. Teachers repeatedly claimed that the bonuses attached to ProComp incentives are not large enough to significantly impact their employment choices.

Focus group participants consistently endorsed a perspective about ProComp incentives that appears to be somewhat at odds with the empirical fact that the average ProComp participant earned more than $19,700 in cumulative incentives over the three-year period of our study, or nearly $6,600 per year. By comparison, in the POINT experiment, annual awards available to teachers as a function of student achievement gains ranged from $5,000 to $15,000. In other words, while the annual salary gains for the average ProComp teacher were not as high as the top range of awards offered to teachers participating in the POINT experiment, they were significant, and in fact, a small number of ProComp teachers did earn bonuses of $15,000 per year or more. These amounts may still not be enough to motivate increased effort if teachers are already working as hard as they can, but they hardly appear to be insulting.

A plausible answer to this riddle comes from Kahneman & Tversky's (1979; 1984) critique of rational choice theory and their introduction of prospect theory as a competing behavioral theory for how people make choices in the presence of risk and uncertainty. In a nutshell, if teachers in a PFP system do not frame their salary bonuses the way that the designers of the system think they should be framed, and if they fail to perceive a strong link between

effort and outcomes, they are unlikely to respond in a manner that is consistent with either the motivation or selection premise. We return to this explanation and its connection to prospect theory shortly after introducing a new way of looking at salary gain, based on *marginal* rather than *cumulative* salary changes.

ProComp Awards From a Different Frame of Reference: Marginal Salary Gains

From a district leadership perspective, the frame of reference against which ProComp should be compared is each teacher's base salary in the absence of any pay incentives. When understood in this sense, any year in which a DPS teacher earns one or more ProComp incentives (see Table 1) represents a year in which that teacher receives a higher salary than he or she would have received in the absence of ProComp. However, from a teacher's perspective, the frame of reference for evaluating the bonuses earned in any given year is very likely to be the *marginal* change in total salary (base salary plus all incentives) from the year before. In other words, teachers effectively reset their perceived "base" salary each year, so that only a new salary that is higher than the prior year salary would be regarded as a gain in salary.

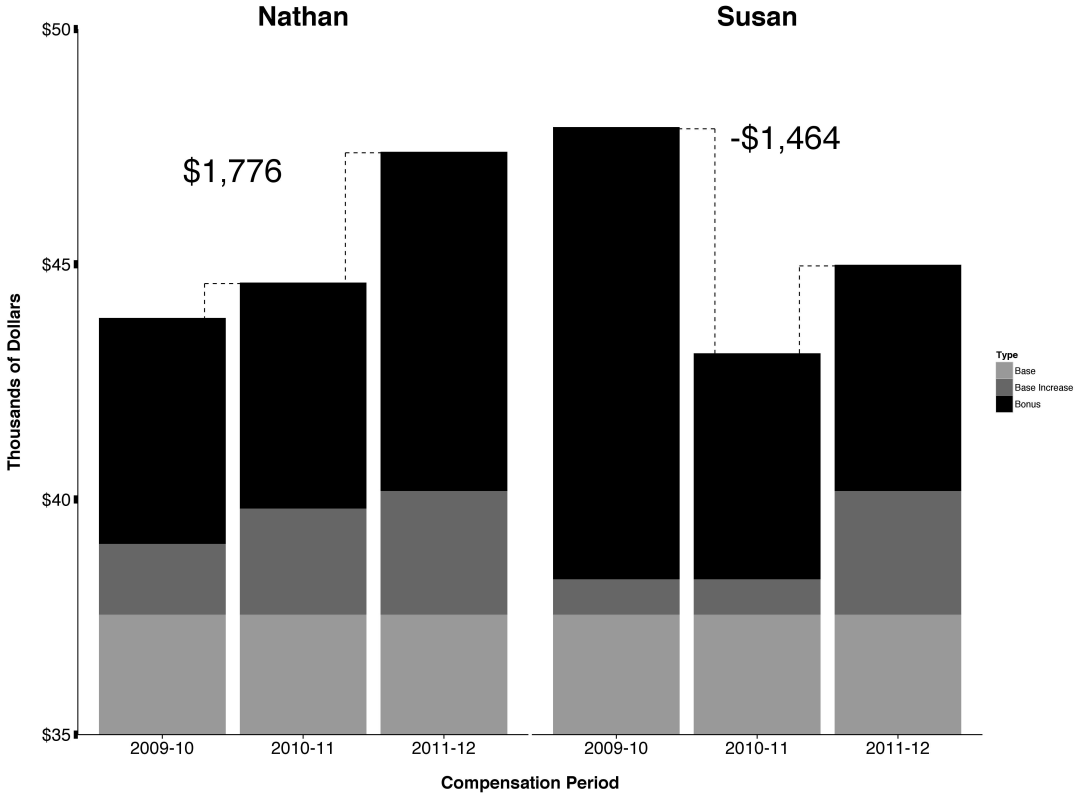

Figure 1. Cumulative and Marginal Salary Gains for Two Teachers

An average marginal gain variable was computed for each teacher by taking the average of two simple differences: the difference in total salaries from 2009-10 to 2010-11, and the difference from 2010-11 to 2011-12. To better understand the distinction between cumulative and marginal salary gains, consider the cases of two young elementary school teachers, Nathan and Susan[1]. At the beginning of the three-year period in our study, both Nathan and Susan were entering their second year with DPS, and both of them held only bachelor's degrees, giving them a starting base salary (not including any rewards for meeting ProComp incentives) of $37,551, shown in light gray in Figure 1. To calculate each teacher's cumulative salary gain, we first "permanently" added the monetary awards associated with any base-building incentives to both

[1] Although the cases represent actual teachers for whom we have data, the names are pseudonyms.

Nathan and Susan's salaries. For example, Nathan earned a $376 reward in 2009-10 for receiving a satisfactory rating on his Comprehensive Professional Evaluation; because this incentive was base-building, it was also added to his salary in all subsequent years (shown in dark gray in Figure 1). Finally, we added one-time bonuses to each teacher's total pay (shown in black). It follows that total salary increase over the three year period for each teacher is the sum of increases to base salary (dark gray) and one-time bonuses (black). By this measure, the two teachers earned similar cumulative gains of more than $23,000 over the three-year period. Note that both of these teachers earned slightly more than the district-wide average of $19,713.

The two teachers look similar on the basis of their cumulative gains over three years, so one might expect them to be equally satisfied with their additional earnings attributable to ProComp participation. However, if we examine their marginal gains over the same period, a different picture emerges. Because Nathan's total salary increased over the entire period, his two marginal gains were positive throughout: from 2009-10 to 2010-11, his gain was $752, and from 2010-11 to 2011-12, his gain was $2,779. His average marginal gain over this time period is $1,776. Susan, on the other hand, actually earned more in 2009-10 than she did in subsequent years. As a result, Susan's two marginal gains were −$4,806 from 2009-10 to 2010-11 and $1,880 from 2010-11 to 2011-12, for an average of −$1,463. Because of this, even though Susan's cumulative salary gain is almost exactly the same as Nathan's over the three-year period, if prior-year salary serves as her frame of reference, Susan would perceive far greater instability and uncertainty in her income.

Table 2. Cumulative vs. Marginal Teacher Salary Gains in ProComp

	Gain Relative to Base Salary in 2009-10			
	Year	ProComp	Traditional	ProComp−Traditional
Cumulative	2009-10	$5,309 ($2,729)	--	--
	2010-11	$6,781 ($2,866)	$725 ($1,148)	$6,056 ($2,963)
	2011-12	$7,623 ($3,022)	$802 ($1,045)	$6,821 ($2,963)
	3-Year Total	$19,713 ($7,266)	$1,527 ($1,897)	$18,186 ($7,240)
	Average	**$6,571 ($2,872)**	**$764 (1,097)**	**$5,807 ($2,963)**
	Gain Relative to Prior Year Salary			
	Year	ProComp	Traditional	ProComp−Traditional
Marginal	09-10 to 10-11	$1,472 ($2,494)	$725 ($1,148)	$747 ($2,593)
	10-11 to 11-12	$843 ($2,343)	$802 ($1,045)	$41 ($2,386)
	Average	**$1,158 ($2,419)**	**$764 ($1,097)**	**$394 ($2,490)**

Note: Values in each cell represent mean and standard deviation, respectively.

For our full population of 2,430 teachers, the distribution of the average marginal salary gain variable is quite different in nature than the average cumulative salary gain variable. These differences are summarized visually in Figure 2, and numerically in the column labeled "ProComp" in Table 2. The average salary gain is $6,571 when 2009-10 base salary is regarded as the frame of reference (cumulative gain); the average salary gain drops to $1,158 when prior year salary is regarded as the frame of reference (marginal gain). If teachers regard prior year salary as their frame of reference for evaluating current year salary, then the average ProComp teacher would perceive a gain in overall pay from one year to the next, but the gain would be perceived as relatively small compared to what is perceived by district leadership. Furthermore, as the bottom panel of Figure 2 indicates, a significant number of teachers are likely to perceive substantial salary losses from one year to the next.

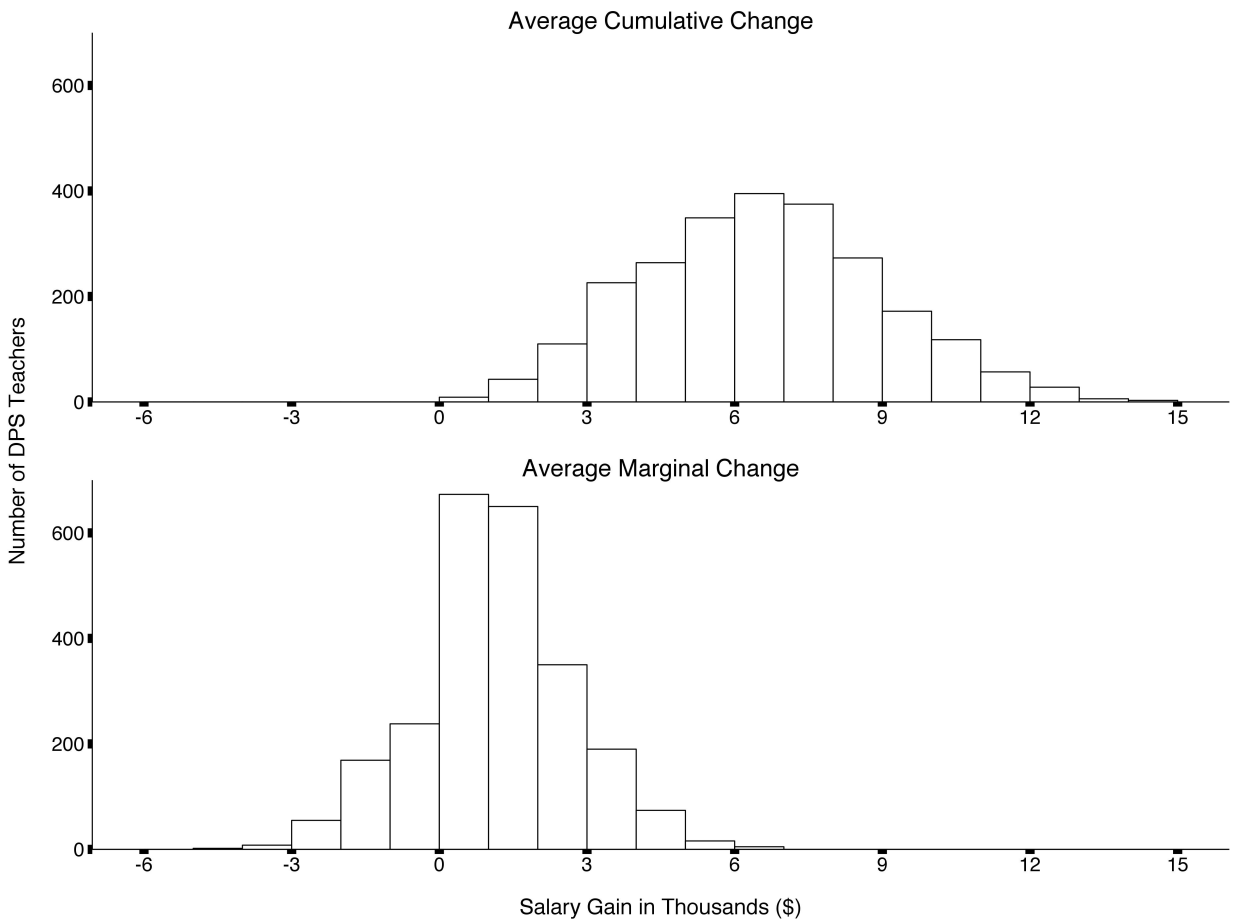

Figure 2. Comparing Average Cumulative and Marginal Salary Gains: 2009-10 to 2011-12

Taking this one step further, we can compare the salary gains for ProComp participants to the gains that would have been realized under the traditional salary schedule that preceded ProComp. The column in Table 2 labeled "Traditional" shows the average amount by which teacher salaries would increase had teachers remained on the traditional schedule, in which increases are tied directly to years of experience and education. Under the traditional salary schedule, the mean salary gain, $764, is smaller, but so is the standard deviation. This is because the distribution of salary increases under the traditional schedule is skewed toward small values with the exception of the subset of teachers earning advanced educational degrees. As a result, the median gain under the traditional schedule is just $221. The advantage of the traditional

schedule is that it is truncated at 0—there are no teachers whose salary in one year will be less than their salary in the previous year. The last column in Table 2, labeled "ProComp−Traditional" shows why district leadership and teachers are likely to have such different perspectives on the impact of ProComp on teacher pay. When taking the cumulative perspective on salary gains, the average annual ProComp effect is $5,807; when taking the marginal perspective on salary gains, the average ProComp effect is just $394.

Prospect Theory

We have presented two competing perspectives that Denver teachers may adopt regarding the monetary awards they earn for meeting ProComp incentives. The frame of reference in the cumulative gain perspective is a base salary that stays constant; the frame of reference in the marginal gain perspective is a moving target. In the cumulative perspective, teachers would take a long-term view on the awards they earn, recognizing that although they make not be equally successful every year at earning incentives, over time the *accumulated* assets due to incentives earned will be greater than they would have been in the absence of ProComp. This perspective would be consistent with rational choice theory (Becker, 1976), in which the choices of self-interested individuals are guided by rational calculations, motivated by the rewards and costs of individual actions. As part of a clever program of experimental research, Kahneman and Tversky (1979; 1984) identified numerous scenarios in which the basic tenets of rational choice theory are violated, and in the process developed a competing behavioral model for decision-making known as prospect theory. A key insight that seems highly relevant to the

perspective being taken by DPS teachers was expressed by Kahneman and Tversky in their 1984 article:

> People do not normally think of relatively small outcomes in terms of states of wealth but rather in terms of gains, losses, and neutral outcomes (such as the maintenance of the status quo). If the effective carriers of subjective value are changes of wealth rather than ultimate states of wealth, as we propose, the psychophysical analysis of outcomes should be applied to gains and losses rather than to total assets. This assumption plays a central role in a treatment of risky choice that we called prospect theory. (Kahneman & Tversky, 1984, p. 342).

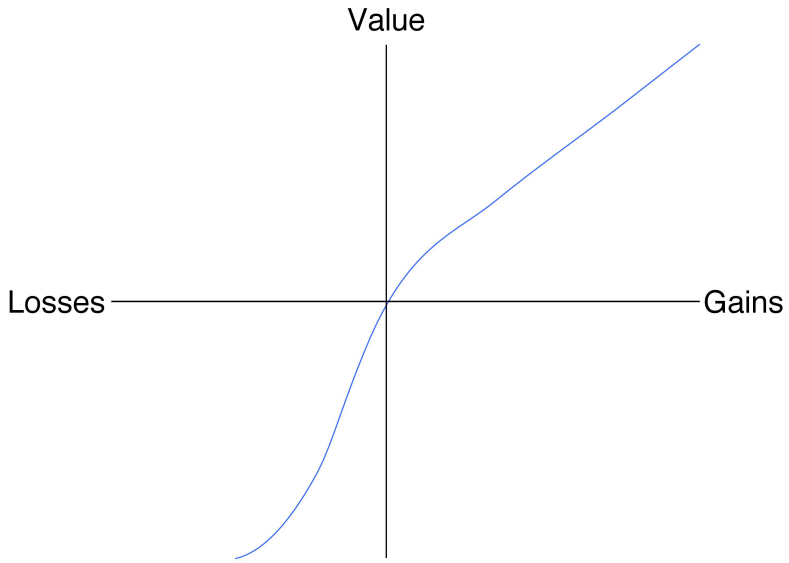

Figure 3. Kahneman & Tversky's Hypothetical Value Function Under Prospect Theory

Two key components of prospect theory can best be understood through visual illustration in Figure 3, which depicts a hypothetical value function for a teacher participating in a PFP salary plan. The first component can be understood by observing that in this figure, a teacher places value (vertical axis) on salary *gains and losses* (horizontal axis), not on total

wealth. The second component can be understood by observing that the curve in this figure is asymmetric as a function of gains and losses. In other words, the decrease in value associated with a small loss is considerably larger than the increase in value associated with a gain of the same size. The shape of the curve for losses can be explained by the fact that people tend to be loss-averse—that is, the loss of $100 is psychologically more aversive than the gain of $100. The first of these components of prospect theory seems consistent with DPS teachers adopting the marginal gain perspective on ProComp awards; the second would explain why teachers would consider the marginal gains they earn in any given year as "insulting," because small gains are perceived to have little value while the negative value associated with even small losses is magnified.

The difference between rational expectations theory and prospect theory can be crystalized by imagining how a teacher would respond if given the choice between the following two options:

- Option 1: Accept a gamble every year in which there is 65% chance of having $2500 added to your salary, and a 35% chance of having $2500 deducted from your salary.
- Option 2: Have $250 added to your salary every year for sure.

According to rational expectations theory, a teacher would be expected to choose Option 1, because the expected value of the gamble is .65*$2500 + .35*(−$2500) = $750. This expected value is three times as large as the amount by which a salary would increase under Option 2. Yet Kahneman and Tversky's research on prospect theory suggests that many—if not most—teachers would prefer Option 2. The values chosen for this hypothetical example are not arbitrary; the highest reward available for meeting a ProComp incentive is close to $2500, and there are a substantial number of teachers who earn these high-reward incentives one year, and

then lose them the next. The expected value of $750 is the average increase attributable to ProComp incentives computed on the basis of the average of marginal salary gains, and the bonus in Option 2 is based on the median increase that would be observed on the basis of the traditional salary schedule (see Table 2).

Discussion

In a well-designed PFP system, one would hope that variations in incentive earnings would depend primarily upon the effort expended by participants. As this relationship between effort and incentives is strengthened, teachers will perceive a higher probability of earning incentives and a lower probability of experiencing a loss from one year to the next. This appears to have been the case in Lazear's analysis of the PFP system implemented at the Safelite Autoglass Company. In windshield installation, where there is a clear and direct connection between effort and outcomes, both the motivation and the selection premises of PFP seem to hold. In contrast, teachers are just one factor that contributes to student learning, and standardized tests only measure limited aspects of this learning. As a consequence, teachers who always exert maximal effort may earn a desirable incentive one year, but then fail to earn it the next for reasons outside their control.

If it were possible to increase the probability of earning an incentive as a function of teacher effort and the true quality of instruction, PFP would be more likely to produce effects on student achievement, and more teachers would be favorably inclined to the system. However, probabilities attached to gains are unlikely to be high in the context of any incentives are attached to average growth in student achievement, because it is well-known that estimates of

teacher impact on student growth suffer from low reliability. Estimates of correlations between average student growth attributed to a teacher from one year to the next range from a low of about 0.20 to a high of about 0.50 (McCaffrey, Sass, Lockwood & Mihaly, 2009). Correlations in this range can be interpreted as indicating that more than half of the variability in student growth across teachers is attributable to measurement error. The implication of this is that the probability associated with earning a salary gain of $2500 may only be a little bit higher than a coin flip. If the PFP theory of action requires teachers to see a strong connection between their pay and their level of productivity, then neither the motivation nor the selection premises will hold if an incentive is strongly influenced by factors outside of a teacher's control, and this may help explain ProComp's lack of impact on student outcomes.

Of course, ProComp is just one PFP system, and it features a diverse array of incentives. Fryer, Levitt, List & Sadoff (2012) crafted an incentive plan that not only anticipates that teachers are more averse to losses than they are to gains, but attempts to capitalize on this response. In their field experiment with teachers in Chicago, a treatment condition consisted of teachers receiving a monetary bonus up front that would have to be returned if the teacher's students did not reach performance targets on standardized tests. Fryer et al. find evidence that teachers exposed to this "loss averse" condition had students with greater increases in achievement than those exposed to more typical incentive conditions in which pay is received after the incentive has been reached. However, this study is incomplete, because it only provides evidence for salary gain or loss in a single year, rather than tracking outcomes across years once teachers have had time to get a sense for the probability of successfully meeting student performance targets. In addition, it seems questionable that leveraging teacher anxiety around loss aversion will produce a sustainable system without a concomitant increase in coaching

(Koretz & Hamilton, 2006), cheating (Jacob & Levitt, 2003), and job dissatisfaction in the long-run.

The bottom line is that the designers of pay for performance systems will need to carefully consider the implications of incentive structures given the anticipated behavioral responses that follow from prospect theory. When teachers are faced with decisions about allocations of their time and energy in the presence of uncertainty about annual salary bonuses, it is predictable that they will focus upon marginal gains and losses, and their perception of the probability of these gains and losses. If, given the expenditure of maximum effort, the probability and magnitude of salary gains are not asymmetrically higher than the probability and magnitude of salary losses, a long-term consequence of a demoralized teaching force—as well as minimal impact on student growth—seems entirely predictable.

References

Atkinson, A., Burgess, S., Gregg, P., Propper, C., Slater, H., & Wilson, D. (2009). Evaluating the impact of performance-related pay for teachers in England. *Labour Economics*, *16*(3), 251–261.

Becker, G. S. 1976. The Economic Approach to Human Behaviour. Chicago: University of Chicago Press.

Besharov, D. J., Podgursky, M. J., & Springer, M. G. (2007). Policy Retrospectives Teacher Performance Pay : A Review, *26*(4), 909–949.

Betebenner, D. (2009). Norm-and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, *28*(4), 42-51.

Figlio, D. N., & Kenny, L. W. (2007). Individual teacher incentives and student performance. *Journal of Public Economics*, *91*(5-6), 901–914. doi:10.1016/j.jpubeco.2006.10.001

Fryer, R., Levitt, S., List, J., & Sadoff, S. (2012). Enhancing the efficacy of teacher incentives through loss aversion: a field experiment. NBER Working Paper 18237 http://www.nber.org/papers/w18237.

Fulbeck, E. S. (2013). Teacher Mobility and Financial Incentives: A Descriptive Analysis of Denver's ProComp. *Educational Evaluation and Policy Analysis*.

Glazerman, S., & Seifullah, A. (2012). An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after Four Years. Final Report. *Mathematica Policy Research, Inc.*

Goldhaber, D., & Hansen, M. (2009). National Board Certification and Teachers' Career Paths: Does NBPTS Certification Influence how Long Teachers Remain in the Profession and Where They Teach? *Education*, *4*(3), 229–262.

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.

Goodman, S. F., & Turner, L. J. (2013). The design of teacher incentive pay and educational outcomes: Evidence from the New York City bonus program. *Journal of Labor Economics*, *31*(2), 409-420.

Hanushek, E. A., & Raymond, M. E. (2006). School accountability and student performance. *Regional Economic Development*, *2*(1), 51–61.

Hanushek, E., & Woessmann, L. (2011). Overview of the symposium on performance pay for teachers. *Economics of Education Review*, *30*(3), 391–393.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263-291.

Kahneman, D., & Tversky, A. (1984). Choices, values and frames. *American Psychologist*, 39(4), 341-350.

Lazear, E. P. (2000). Performance Pay and Productivity. *The American Economic Review*, *90*(5), 1346–1361. Retrieved from http://www.nber.org/papers/w5672

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

McCaffrey, D. F, Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, *4*(4), 572–606.

Muralidharan, K., & Sundararaman, V. (2009). *Teacher Performance Pay: Experimental Evidence from India* (No. 15323).

Podgursky, M., & Springer, M. (2011). Teacher compensation systems in the United States K-12 public school system. *National Tax Journal*, *64*(1), 165–192.

Podgursky, M., & Springer, M. G. (2007). Credentials Versus Performance: Review of the Teacher Performance Pay Research. *Peabody Journal of Education*, *82*(4), 551–573.

Springer, M. G., Ballou, D., Hamilton, L., Le, V. N., Lockwood, J. R., McCaffrey, D., Pepper, M, & Stecher, B. (2010). *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT)*. Nashville, TN: National Center on Performance Incentives at Vanderbilt University.

Springer, M. G., & Winters, M. A. (2009). The NYC Teacher Pay-for-Performance Program: Early Evidence from a Randomized Trial. Civic Report No. 56. *Manhattan Institute for Policy Research*.

Woessmann, L. (2011). Cross-country evidence on teacher performance pay. *Economics of Education Review*, *30*(3), 404–418.