Strategic Pay Reform: A Student Outcomes-Based Evaluation of Denver’s ProComp Teacher Pay Initiative*

CEDR Working Paper 2011-3

Dan Goldhaber & Joe Walch
Center for Education Data & Research
University of Washington Bothell

* The research presented here utilizes confidential data from Denver Public Schools. We gratefully acknowledge the receipt of these data. We also wish to thank Wesley Bignell for research assistance and John Tyler and Ed Wiley for helpful comments on early findings. The views expressed in this working paper are solely those of the authors and do not necessarily represent the Center for Education Data & Research or the University of Washington. Any and all errors are the responsibility of the authors.

2011 © Center for Education Data & Research.
I. Growing Interest in Strategic Compensation Reform

There is significant and growing interest in teacher pay reform as a number of states and localities have begun experimenting with departures from the single salary schedule—a pay system employed in most school districts, which links teacher pay solely to degree and experience level (Chait, 2007; Goldhaber, 2009; Podgursky and Springer, 2007). Florida, Minnesota, and Texas, for example, have all embarked on high-profile pay experiments that include performance pay, arguably the most controversial type of pay reform, as a central component. These states are joined by urban school systems such as Denver, Houston, and New York City that have launched reform initiatives. The federal government is encouraging pay reform through its Teacher Incentive Fund (TIF) and, more recently, through the Race to the Top initiative.1

There are good reasons to focus on teacher pay as an avenue for school reform. We know that the quality of teachers is the most important schooling attribute explaining student achievement (Rivkin et al., 2005), and the investment in teachers is typically the single largest share of local school system budgets; teacher salaries usually comprise 50-60 percent of total spending (Monk et al., 1997; Rothstein and Miles, 1995; Digest of Education Statistics 2010). Moreover, education advocates, analysts, and policymakers routinely criticize typical teacher compensation systems for being too uniform, rigid, and only weakly related to effectiveness.2 In particular, they argue that these systems make teaching relatively undesirable for highly productive people and/or those with technical skills.3

1 Note that we say teacher “pay” as opposed to teacher “compensation,” which in addition to pay would also include benefits.
2 While interest in teacher pay reform may be on the rise, calls for reform are certainly not new. Pay for performance (PFP) was, for example, one of the recommendations of the 1983 A Nation at Risk Report. And, numerous school districts adopted, and then abandoned, merit plans throughout the 1980s (Hatry et al., 1994; Murnane and Cohen, 1986).
3 Research has also shown that there is a significant amount of variation in the effectiveness of in-service teachers (Aaronson et al., 2007; Goldhaber and Hansen, 2010; Rivkin et al., 2005); the difference between students having a very effective and very ineffective teacher is estimated to be equivalent to more than a full grade level of learning (Hanushek, 1992).
4 The observed “value-added” variation in teacher effectiveness is only weakly (at best) linked to the determinants of teacher compensation under a single salary schedule. There is clear evidence that teachers become more effective in their first few years in the classroom, but it also shows that this relationship levels out beyond the first three to six years of experience (Clotfelter et al., 2006). But while experience does predict teacher effectiveness, there is still a significant share of, for instance, novice teachers that are more effective than third year teachers and vice versa. See, for instance, Figure 4 in Gordon et al. (2006). Teacher degree level appears to be completely unrelated to effectiveness outside of a few narrow exceptions. For a more comprehensive discussion, see Goldhaber and Brewer (1997) and Hanushek (1986, 1997).
5 It is certainly true that the structure of compensation in education is very different from the way the broader labor market tends to function (Eide et al., 2004). As a whole, private sector compensation generally reflects not only individual attributes such as cognitive or technical skills, but also performance on the job (Bretz and Milkovich, 1989; Brewer et al., 1999; Grogger and Eide, 1995; Murnane et al., 1995). And, there is some evidence that the divergence in compensation structure in and outside of the teaching profession may encourage the loss of highly productive teachers (e.g. West and Chingos, forthcoming) and explain a long-term decline in the average academic caliber—measured by standardized test scores and/or the selectivity of colleges—of teachers (Corcoran et al., 2004; Hanushek and Pace, 1995; Hoxby and Leigh, 2005). Finally, the prevailing teacher pay system typically does not recognize the difficulty of teaching assignment, which may contribute to the inequitable distribution of teachers across students (Boyd et al., 2009). In short, teacher pay reform advocates suggest that, “teachers who do well, those
On the flipside, the single salary schedule has straightforward advantages: everyone is rewarded equally based on objective criteria; the system is predictable and easy to understand. Bacharach et al. (1984), for instance, note that salary schedules have particular advantages in schools: they reflect the fact that teachers learn from experience, they allow teachers to have high expectations for their students without worrying that it threatens their income security, they avoid competition between teachers that might inhibit collaboration or knowledge exchanges, they avoid the perception that students are a barrier to pay increases, and they allow administrators to hold more holistic expectations about teacher performance (rather than focus only on what is easily measured).

Much of the headline-grabbing attention of pay reform focuses on pay for performance (PFP). The empirical evidence of performance pay’s effectiveness is mixed. Some research suggests a connection between pay for performance and student achievement (e.g. Figlio and Kenny, 2007), but recent evidence from an experiment testing this notion suggests that PFP does not lead to better student outcomes, at least through increased effort of incumbent teachers (Springer et al., 2010).6

In a study of teacher performance incentives in Chicago using experimental and quasi-experimental research designs, Glazerman and Seifullah (2010) failed to find any evidence that financial incentives lead to improvement in student test scores or increased teacher retention. Another study of the same program (Springer et al., 2008) focused specifically on math scores and came to similar conclusions for middle and high school teachers, though they did find evidence of increased student test scores for elementary school teachers.

Denver’s Professional Compensation System for Teachers (ProComp) represents what is arguably the nation’s most high-profile example of strategic pay reform.7 ProComp’s origins date back to 1999 when Denver Public Schools (DPS), with the local teachers union’s cooperation, initiated a pilot program meant to increase student achievement and attract and retain highly effective teachers. The program gained support as a pilot, and in 2004, union members ratified a proposal for a more comprehensive program which became ProComp. In 2005, Denver voters approved a tax hike to provide funding of $25 million for the program, and an additional $22.6 million was provided by a TIF grant.

The 2004 agreement between the teachers union and the district required that teachers hired after January 1, 2006 follow the ProComp salary schedule, while incumbent teachers could opt in to the program or stay with the traditional salary schedule. ProComp participants have the

---


7 The concept of strategic pay reform (SPR) is really much broader, encompassing not just the notion that pay should be connected to performance, but also that pay might be used to encourage a more equitable distribution of teachers within districts, reduce the “churn” of teachers, lead to stronger connections between teacher pay, evaluation, and professional development systems, and enhance system-wide learning (e.g. teachers learning about effective practice from each other). All of these broader concepts are consistent with the ultimate objective of increasing student achievement, particularly for disadvantaged students.
opportunity to earn bonuses by meeting requirements in four areas: 1) knowledge and skills, 2) comprehensive professional evaluation, 3) market incentives, and 4) student growth.⁸

In order to assess whether ProComp has succeeded in increasing student achievement, we explore three primary questions:

- Is there a ProComp “system effect”? Is student achievement higher in years after the implementation of ProComp?⁹
- How effective are teachers who choose to opt in to ProComp, compared to teachers who choose not to opt in?
- How does the allocation of rewards in the ProComp system correspond to teacher effectiveness?

II. Reward Structure Under ProComp

Under ProComp teachers have the opportunity to earn bonuses by meeting requirements in four broad areas:¹⁰

1. Knowledge and Skills: Professional development units, advanced degrees and licenses, and tuition and student loan reimbursement;
2. Comprehensive Professional Evaluation;
4. Student Growth: Student growth objectives, exceeding expectations on the Colorado Student Assessment Program (CSAP) tests, top performing schools, and high-growth schools.

The bonuses that are paid out are relative to an index salary. The 2010 index salary was $37,551, so a 2 percent bonus amounts to an additional $751, or total salary of $38,302.

In the first area, “Knowledge and Skills” (K&S) teachers can earn a 9 percent bonus for each advanced degree and professional license and a 2 percent bonus for participating in a

---

⁸ The payment bonus system is discussed in more detail in Section II of this report.
⁹ As we describe in more detail below, we are unable to separate the ProComp “system effect” from the effect of other factors associated with the ProComp time period.
¹⁰ In 2008, the Denver Classroom Teachers Association voted to amend the ProComp agreement, increasing the base salary, adding the student loan reimbursement option, and increasing the bonus amount for the following incentives: hard to serve schools, hard to staff assignments, top performing schools, high growth schools, exceeds expectations, and tuition and student loan maximums. The amounts for bonuses reported in this section are updated to reflect the changes made in 2008.
professional development program. Teachers can also receive up to $1,000 per year (maximum of $4,000 total) for tuition and student loan reimbursement.

Under the “Comprehensive Professional Evaluation” (CPE) area probationary teachers can earn a 1 percent bonus each year and non-probationary teachers can earn a 3 percent bonus every three years for a successful evaluation.

The “Market Incentives” (MI) area rewards teachers for teaching in “hard-to-serve” schools and “hard-to-staff” assignments. Alternative schools and schools with a high percentage of students on free and reduced-priced lunch (FRL) are identified as “hard-to-serve,” and ProComp teachers in these schools are eligible for a 6.4 percent bonus. Teachers can also earn a 6.4 percent bonus for teaching in “hard-to-staff” assignments, defined as those teaching positions with high turnover and vacancy.

Finally, under the “Student Growth” area, teachers can earn a 6.4 percent bonus for teaching in a “top performing school” and a 6.4 percent bonus for teaching in a “high growth school.” Schools are designated as top performing and high growth based on the School Performance Framework (SPF), an evaluation instrument developed by DPS to assess school performance in terms of student achievement. Teachers whose students exceed expectations on the Colorado State Assessment Program (CSAP) tests can earn a 6.4 percent bonus. The “exceeds expectations” (EE) bonus is based on students’ conditional growth percentiles, which compares each student with a group of students with a similar CSAP score history. Students exceed DPS’ CSAP expectations if they score in the top 55 percent of their comparison group, and teachers earn the 6.4 percent bonus if at least half of their class performs well enough to exceed CSAP expectations.

Teachers also have the opportunity to earn a 1 percent bonus for each “Student Growth Objective,” (SGO) with a maximum of 2 percent each year. Student growth objectives are goals set individually by each teacher and approved by the principal. Student growth objectives vary from teacher to teacher but usually involve improvement on a classroom assessment, and they cannot be based on CSAP scores.

There is considerable variation in the proportion of teachers who receive each type of incentive. For example, a consistently high percentage, ranging from about 50 to 65 percent, of teachers between 2006-07 and 2009-10 receive at least some portion of the K&S incentive, mainly as a result of completing PDUs. Likewise, 70 to 80 percent of teachers receive an SGO

---

11 Teachers can receive payment for one Professional Development Unit (PDU) each year, but can “bank” them and save them for later years if they have more than one for a particular year.
12 In general, teachers are considered “probationary” during the first three years in the district.
13 In 2010-2011, elementary schools with at least 87 percent of students FRL eligible, middle schools with at least 85 percent of students FRL eligible, and high schools with 75 percent of students FRL eligible were considered “hard-to-serve” schools.
14 Hard-to-staff designations are reviewed every year and are based on several data sources, including national, state, and district-wide data on the supply of teachers.
15 The SPF takes into account a broad range of measures, including CSAP and AYP performance and improvement from previous years, and measures progress with respect to other schools with similar demographic characteristics.
incentive. A far lower percentage of teachers receive MI (35 to 65 percent) or EE (5 to 14 percent) incentives during this time.

Teachers employed by DPS prior to January 2006 had the opportunity to opt in to the ProComp system on a voluntary basis, but those hired after were automatically enrolled. Figure 2 shows the percentage of the DPS workforce that is enrolled in ProComp.16

### III. Analytic Approach and Data

There are at least three distinct causal pathways through which ProComp could lead to better student achievement (see Figure 3). First, we would expect teachers who are being rewarded for student achievement gains (on Colorado’s state assessments, CSAP tests) to focus their instruction and effort around student achievement on these tests. This first pathway then is really about individual teacher productivity at any point in time. Wiley et al. (2010), who investigate the effects of ProComp, find little evidence of this type of teacher effort-related productivity effect.17

Beyond changes to teacher effort at a point in time, we might expect that ProComp—to the degree that it improves the feedback and professional development teachers receive—should make them more productive over time. This second pathway may result from teachers learning more (under ProComp) about areas of weakness and addressing them through professional development. But the new system may also lead teachers to learn more from each other through less formalized channels. In particular, we might expect role-modeling effects that occur precisely because ProComp identifies and rewards successful teacher excellence in a way that the old DPS compensation system did not.18

The last pathway involves compositional changes to the DPS teacher workforce—that is, the potential that ProComp influences the pipeline of teachers into and out of the district. Given that ProComp is designed to reward teacher success, we might expect the pay system to attract teachers who are likely to be successful under the ProComp system; consequently, we would expect ProComp to have impacts on both the number and quality of teacher applicants (see Chapter 9 for more detail). Similarly, we would expect that teachers who are successful under ProComp would be relatively likely to remain in the district and unsuccessful teachers would be relatively likely to leave. Wiley et al. (2010) finds some evidence of a ProComp composition effect in that teachers hired after the implementation of ProComp are found to be more effective than those with similar years of experience hired prior to the adoption of the new pay system.

---

16 As laid out in the ProComp agreement between DCTA and DPS, once teachers opt into ProComp they cannot opt out. However, there was a one-time exception included in the 2008-2009 contract that allowed teachers who previously voluntarily opted in to ProComp to opt out of the program in the 2008-2009 school year. Approximately 40 teachers in the dataset chose to opt out.

17 This is consistent with the literature cited above (Springer et al., 2010) on pay for performance.

18 Jackson and Bruegmann (2009) find that teachers perform better when they have more effective colleagues and that these effects are strongest for less experienced teachers. The results suggest that teachers’ productivity is influenced through learning and knowledge spillovers that occur among peers.
We estimate a variety of analytic models designed to assess the effect of ProComp on the student achievement, with an eye toward trying to determine whether ProComp is impacting teacher effectiveness through any of the pathways described above.

A. Is There a ProComp “System Effect”?

In order to test the presence of a system effect, we begin with simple pre-post models to compare the student achievement from the time period before ProComp was implemented to student achievement during the ProComp years. We estimate the ProComp system effects with the following equation:

$$ A_{ijt} = \beta_{procompyears} + \gamma X_{it} + \delta T_{jt} + \epsilon_{ijt} \quad (1) $$

In this model, the test achievement of student $i$, taught by teacher $j$, in year $t$, $A_{ijt}$, is regressed on the ProComp time period (with pre-ProComp as the omitted category), student background characteristics, $X_{it}$ (prior year CSAP scores in math and reading, race/ethnicity, learning disability, free/reduced price lunch status, and grade), and teacher characteristics, $T_{jt}$, (degree level, experience, and experience squared). We also estimate these models separately by grade configuration.

We also investigate the effect of the significant changes made to the ProComp system in 2008 (details in footnote 9). We designate the years after these changes were implemented as “ProComp 2.0” and compare the student achievement during the pre-ProComp (2002-03 through 2004-05), ProComp 1.0 (2005-06 through 2007-08), and ProComp 2.0 (2008-09 through 2009-10) time periods. Indicators for ProComp1.0 and ProComp2.0 are included in this second model.

The main coefficient of interest in terms of the ProComp system is $\beta$, as it identifies differences in student achievement associated with implementation of ProComp (we also assess different phases of ProComp) as compared to achievement in the pre-ProComp time period. In (1) above, identification of what we are calling the “ProComp effect” is based solely on differences in student achievement over time. It is important to note that the ProComp system should be considered as broader than just changes to the pay structure. For instance, DPS revamped the teacher evaluation system that applies to all teachers (see Chapter 6 and 2004 ProComp Agreement for more detail). DPS also upgraded their information systems to allow teachers and administrators more timely access to student and teacher data (see Chapter 3 for more detail).

Given this, ProComp system findings would not necessarily be associated solely with the new salary structures. Additionally, we cannot distinguish between effects that are based solely on the adoption of the ProComp system and other factors concurrent with the implementation of ProComp 1.0 or 2.0 (e.g. a new student health initiative). Finally model (1) does not necessarily
It is potentially important to distinguish between workforce composition and individual teacher productivity effects. In particular, were the case that a ProComp effect is attributable solely to workforce effects, there is the potential that, in the short-run, gains or losses of teacher human capital from DPS are partially offset by losses or gains in human capital from other school districts in or outside of Colorado. \(^{20}\) Gains associated with the productivity of individual teachers on the other hand suggest that the capacity for change amongst incumbent teachers, which offers the potential for both more immediate improvements in the quality of the teacher workforce.

The evidence for changes in productivity of incumbent teachers is mixed. As we discussed above, the rigorous evidence of the impact of pay-for-performance systems on the productivity of individual teachers is not terribly promising (e.g. Springer et al., 2010). There is, however, some new evidence that comprehensive evaluations can impact the productivity of individual teachers. Taylor and Tyler (2011), for instance, find that students in Cincinnati Public Schools score about 10 percent of a standard deviation higher in math if their teacher has participated in the district’s evaluation program, and that this effect persists after the evaluation year. They did not, however, find an effect on student reading scores. As a comprehensive performance evaluation (CPE) is a central part of the ProComp system (see Chapter 6 for more detail), we might expect to find ProComp impacting the performance of individual teachers.

In order to distinguish workforce composition effects from what might be individual teacher productivity changes, we estimate a variant of (1), which includes a teacher fixed effect, \(\tau_j\):

\[
A_{ijt} = \beta_{procomp\text{years}} + \gamma X_{it} + \tau_j + \epsilon_{ijt} \tag{2}
\]

In this specification the ProComp effect is identified based on \textit{within} teacher differences in student achievement over time so does not reflect any influence of ProComp that make work through workforce composition. \(^{21}\) This specification also accounts for any time-invariant, non-random differences in the assignment of students to specific teachers.

\begin{itemize}
\item B. ProComp Enrollment and the Performance of Voluntary Participants
\end{itemize}

\(^{19}\) Theory would predict that teachers who would financial benefit from ProComp’s reward system would be likely to be sorted into the DPS workforce. And, since much of what determines teacher effectiveness is not associated with easily quantifiable variables like degree and experience levels (Goldhaber et al., 1999; Rivkin et al., 2005), this workforce sorting may not be identified with the teacher variables included in the vector T.

\(^{20}\) In the longer-run one might imagine that the workforce effects could influence the distribution of individuals in the teaching profession.

\(^{21}\) The coefficient for the ProComp years variable is informed solely by teachers who are in the data in ProComp years and pre-ProComp years.
The above explores student achievement in Denver during different ProComp periods but it does not explore whether any achievement differential appears to be associated with those teachers actually enrolled in the ProComp system or whether there are differences between those enrolled teachers who voluntarily opted in versus those who were mandatory placements (i.e. hired into DPS after January 2006).

Whether there are productivity gains associated with different types of DPS teachers is conceptually important for at least two reasons. First, as we noted above, while ProComp is explicitly designed as a pay reform, the implementation of the system entailed the development and restructuring of ancillary systems, from human resources to data systems. It is therefore conceivable that the types of information about teacher performance and performance feedback loops could provide benefits to teachers who are not actually enrolled in ProComp’s alternative pay plan. On the other hand, pay reform is often quite controversial (Goldhaber, 2009), and the strife pay systems that differentiate teachers potentially cause could also lead to spillover effects that negatively impact those teachers not enrolled in the system. Knowing whether productivity changes associated with a comprehensive reform like ProComp are concentrated amongst teachers enrolled in the system will help policymakers decide whether to include pay reform when weighing human capital decisions.

Second, one of the goals of the ProComp system is to make the teaching profession more attractive, especially to individuals who would be, or are, effective in the classroom. Thus, it is informative to know what kind of teachers appear to prefer the ProComp system as this will have a long-term influence on any ProComp workforce composition effects (i.e. changes in workforce quality brought about by the effectiveness of teachers who enter or leave DPS). Workforce composition effects are potentially quite important. Lazear (2000), for instance, finds that much of the gains associated from pay for performance in the private sector result from more productive workers sorting into a performance-based system.\textsuperscript{22}

We cannot directly observe the preferences of teachers, but we can indirectly assess the desirability of working in the ProComp system for teachers of varying effectiveness by focusing on differences between teachers who voluntarily opt into the system and those that do not.

We explore these issues utilizing a third model specification that builds off of (2) above by including variables that indicate whether a teacher is a ProComp participant and whether they opted in voluntarily:

\textsuperscript{22} Theory would suggest more effective teachers who have much to gain from a pay for performance system to be more likely to opt in to ProComp because of the potential rewards (Gibbons, 2005). There exists relatively little empirical evidence from U.S. schools on this potentially important labor force composition effect, but Muralidharan and Sundararamin (2011) investigate a PFP program in India and find a positive and statistically significant relationship between teachers’ preferences for performance pay and the average gains of their students (the survey was administered \textit{before} teachers knew those gains). This suggests that effective teachers know who they are and wish to be rewarded for their performance, and, further, that performance-based pay systems may yield workforce productivity benefits associated with more sorting, that is, those with teaching talent opting into the teaching profession and the more effective teachers opting to remain in the profession. Moreover, Woessmann (2011) estimates the effects of teacher pay for performance systems using cross-national data and finds a significant association between these systems and student achievement in math, science, and reading. Using cross-national data helps identify longer term, general-equilibrium effects that better incorporate both the incentive and sorting effects associated with pay for performance systems.
\[ A_{ijt} = \beta_{procomp} \text{years} + \psi_{participant} + \delta_{voluntary} + \gamma_{X_{it}} + \delta_{T_{jt}} + \epsilon_{ijt} \] (3)

We also estimate a variant of (3) that includes teacher fixed effects. In the fixed effects model, the within-teacher effect of voluntarily opting into ProComp is indicated by the ProComp participant coefficient, \( \psi \). The identification for this estimate is based on the 586 teachers (1,274 teacher-years) in the analytic sample who have observations both prior to voluntarily opting into ProComp and after they opt in.

**C. How Effectively Are ProComp Rewards Targeted?**

A central goal of ProComp is to reward teacher effectiveness. We gauge the success of ProComp’s targeting in two steps. In step one we estimate individual teacher effectiveness, and then in step two we use this estimate in models that predict the probability of receiving a particular ProComp reward.

The estimation of teacher effectiveness is complicated by the fact that there is no universally accepted method for calculating a teacher’s value-added contribution to student achievement.\(^{23}\) The primary specification we utilize is:

\[ A_{ijt} = \gamma_{X_{it}} + \tau_{jt} + \epsilon_{ijt} \] (4)

This specification allows for the calculation of estimated teacher-year effects, \( \hat{\tau}_{jt} \).\(^{24}\) However, for some rewards, particularly SGOs, it is not clear whether teachers are judged relative to other teachers in the district or other teachers in a school, so we also estimate a variant of (4) that includes school fixed effects.

In order to account for measurement error in the estimation of the teacher effects, we use an empirical Bayes (EB) adjustment common in the literature (Boyd et al., 2008; McCaffrey et al., 2008). This adjustment weights the estimates based on their reliability, and shrinks unreliable estimates with large standard errors back towards the grand mean of the population.\(^{25}\)

---

\(^{23}\) Research shows that methodology and teaching context (e.g. the sorting of students across classrooms) can influence the measure (Ballou, Sanders, & Wright, 2004; McCaffrey et al., 2004; Rothstein, 2010; Rubin, Stuart, & Zanutto, 2004; Tekwe et al., 2004).

\(^{24}\) Teacher effectiveness estimates are generated using the stata command *felsdvregdm* (Mihaly et al, 2010) with the effectiveness estimates within each year and grade configuration combination summing to 0. Although teachers are only compared to teachers in the same year-grade configuration in the estimation of \( \hat{\tau} \), we can compare each \( \hat{\tau} \) across grade levels. Thus, an elementary teacher with a \( \hat{\tau} \) of .5 is defined as being as effective as a middle school teacher with a \( \hat{\tau} \) of .5.

\(^{25}\) The EB adjusted teacher effectiveness estimates are highly correlated with the unadjusted estimates: \( r = .96 \) or higher for all models.
In step two, we use the various teacher effectiveness estimates to gauge how well the ProComp rewards are targeted towards effective teaching. This is accomplished by estimating (using a logistic regression) the probability of the receipt of a reward:

\[
\text{reward}_{jt} = \eta \hat{\tau}_{jt} + \zeta \text{Year} + \rho (\text{Year} \times \hat{\tau}_{jt}) + \epsilon_{jt} \quad (5)
\]

In (5), the ProComp reward to teacher j (in year t) is regressed against teacher effectiveness estimates, \( \hat{\tau}_{jt} \), year dummies, and teacher effectiveness-year interactions.

Our coefficients of interest are \( \eta \) and \( \rho \). The coefficient \( \eta \) indicates whether more effective teachers are more likely to receive a particular reward and \( \rho \) provides an indicator of whether the targeting of teachers for awards changes over time. One might, for instance, think that the system would become more effective over time at identifying the right (i.e. more effective) teachers to reward. However, it is also possible that teachers learn how to game the system better over time (e.g. Cullen and Reback, 2006; Figlio, 2006; and Jacob, 2005) in order to receive a reward and/or political pressure leads to less effective targeting.

**Analytic Sample**

We use administrative data from DPS from school years 2002-03 to 2009-10 to investigate differences in the achievement of students. In 2010, DPS included approximately 80,000 students and about 4,500 teachers. Key to our investigation is the fact that teachers and students can be linked within each year at the middle and high-school levels, and from the 2005-06 to 2009-10 school years at the elementary level, and teachers can be linked over time.

DPS teacher records include variables such as race/ethnicity, years of experience, degree level, ProComp status, whether the teacher opted-in to ProComp voluntarily, and the rewards earned under ProComp. Student-level records contain information on student background,

---

26 Note that it is possible for a teacher who teaches both math and reading to get a reward based on high performance in one area, even if performance in the other area is low. The data does not differentiate the subject for which a teacher earns a reward.

27 Murnane and Cohen (1986) note that one of the reasons that many early pay for performance systems failed is that performance bonuses were indiscriminately awarded, leading to the financial collapse of these systems.

28 Student and teacher counts are taken from the DPS website: http://communications.dpsk12.org/newsroom/facts-and-figures/about-denver-public-schools/

29 Hence, we cannot compare ProComp years to pre-ProComp years for elementary students and teachers.

30 Data for the pre-ProComp and ProComp years are derived from different sources, and could not be directly linked together due to the fact that different teacher masking systems were used. We obtained a cross-walk between the two systems, but a smaller percentage of teachers than might be expected (e.g. given teacher attrition) were linkable longitudinally. For instance, of the 707 teachers in the pre-ProComp analysis sample, 454 (71 percent) also show up in one or more ProComp years. We cannot determine the reason for the lower than expected linkage rate, but t-tests of the means of select teacher and classroom characteristics show only small (and occasionally statistically significant) differences between those teachers who we can and cannot link longitudinally. These results are available upon request.
including race/ethnicity, learning disability, free/reduced price lunch status, grade, and student achievement on standardized tests.

We use the Colorado Student Assessment Program (CSAP) tests, which are required by Colorado law, and designed to measure students’ mastery of the Colorado Model Content Standards in grades 3–10, as our dependent variable. Because of limitations in the data and the way the CSAP is administered, we choose to focus on the analysis on math and reading, even though students also take the Writing CSAP in grades 3–10 and the Science CSAP in grades 5, 8, and 10. Since our analytic strategy requires a prior-year test score, the “gap years” in the science test make it impossible to attribute the learning gains in science to a particular teacher. Our dataset does not include writing scores prior to the 2005-06 school year, and in school years 2005-06 to 2008-09 only includes writing scores in elementary grades. We normalize all test results within grade and subject, and across years, based on observations from all students in the unrestricted sample.31, 32

DPS, as well as the state of Colorado, uses student growth percentiles generated by the Colorado Growth Model (CGM) to measure student achievement (Report of the Technical Advisory Panel for the Longitudinal Analysis of Student Assessment, 2008), and, in some districts, teacher or school effectiveness. The value-added models we describe in this section are estimated differently from the CGM, but findings on ProComp are similar whether one uses the estimation strategies described here or employs CGM as a metric (see, for instance, Wiley et al., 2010). This is not surprising as individual teacher value-added estimates are highly correlated with the median CGM for each teacher; in our analytic sample we find a correlations between the two of over 0.8 in math and 0.6 in reading (employing models described by equation 4 above).

For our analyses we exclude students who do not have a prior test score and exclude teachers with fewer than 5 students in the data. After the above restrictions, our analytic sample includes 73,197 unique students (163,565 student-year observations), and 2,539 unique teachers (5,820 teacher-year observations). Table 1 shows selected student sample statistics for the unrestricted sample, and restricted samples by their teachers’ ProComp enrollment status.

There are no considerable differences in the proportion of minority students or the proportion of students with a disability between the unrestricted and restricted samples, while students in the restricted sample were slightly more likely to be on free/reduced price lunch. Within the restricted sample, teachers participating in ProComp were slightly more likely to teach minority students than non-ProComp teachers, while non-ProComp teachers were slightly more likely to teach students with free/reduced price lunch and with learning disabilities. Students of ProComp teachers also had higher standardized CSAP Math scores.

31 The CSAP tests are vertically aligned and administered year over year to students in math and reading so, in theory, normalization is not necessary for our analysis. However, the normalization facilitates interpretation as the coefficients can be interpreted as effect sizes (the percentage change in standard deviations on the CSAP test) and more easily compared to other findings in the literature.

32 It is common to also normalize within year, however, such a normalization would preclude the detection of growth in achievement over time.
IV. Results

A. Is there a ProComp “System Effect”?

Before discussing the ProComp findings, a brief discussion of the estimated impact of teacher effectiveness on student achievement is helpful in providing context for interpreting the effects described below. Estimating individual teacher effectiveness (as in equation 4 above), we find a one standard deviation increase in teacher effectiveness (for example, moving from the median to the 84th percentile) to correspond to an increase of .09 to .18 standard deviations in student achievement. These effect sizes vary slightly by grade configuration, subject, and model specification, with somewhat larger effects at the elementary level.33 These estimates are very much in line with published findings elsewhere in the literature on teacher effectiveness (Hanushek and Rivkin, 2010).

Estimates from the literature suggest that the differential between a novice (0 years of prior experience) and second-year teacher (1 year of prior experience) is in the neighborhood of .01-.07 standard deviations of student achievement (Rockoff 2004, Rivkin et. al, 2005). We find similar returns to experience in our models, with larger effect sizes in math (.02-.05) than in reading (0 to .02), and smaller effect sizes in high school compared to elementary and middle schools.

We can also use the standard deviation of the CSAP scores to get a reference point for effect sizes. The CSAP standard deviation in our analytic sample is about 80 points in both math and reading, across grade configurations.34 This means that an effect size of .05 translates to an increase in CSAP scores by four points.

Table 2 shows the findings for student achievement models at the elementary level. The results for control variables in the models (e.g. base year test scores, student demographics, etc.) are consistent across grade configuration and model specifications, and consistent with the literature (e.g. Rivkin et al., 2005). For instance, minority students (African-American and Hispanic), participants in the free and reduced-price lunch program, and/or those who have learning disabilities score lower, all else equal, than students who are white, nonparticipants in the free and reduced-price lunch program, and/or do not have learning disabilities. There is little consistency on the findings on teacher degree and experience level across grade levels; again, this is consistent with the literature.

In general we find evidence suggesting the presence of ProComp system effects, though recall that we cannot definitively ascribe this to the ProComp system, or the alternative compensation aspects of ProComp. In particular, the estimation strategy we employ makes it

---

33 Empirical Bayes shrunken estimates of a one standard deviation change in teacher effectiveness in math are .18 at the elementary level, .14 at the middle school level, and .09 at the high school level; the corresponding estimates in reading are .13, .16, and .15. The effect sizes for the within-school teacher effectiveness estimates are slightly smaller: in math they are .15 at the elementary level, .13 at the middle school level, and .08 at the high school level; and in reading the corresponding estimates are .09, .10, and .09.

34 The standard deviations by grade and year in the analytic sample are very similar to those reported by CDE (CDE website: http://elm.cde.state.co.us/cognos8bi/cgi-bin/cognos.cgi).
impossible to separate ProComp system effects from any other unobserved factors that affect student achievement that are concurrent with ProComp implementation.

The elementary results are reported separately from the middle and high school results because the lack of student-teacher linked data at the elementary level in pre-ProComp years means we can only compare ProComp 1.0 (school years 2005-06 to 2007-08) to ProComp 2.0 (school years 2008-09 and 2009-10). In these models, the omitted reference category is the ProComp 1.0 years; the first two columns show the results for math and the next two for reading.

The results with teacher covariates suggest small increases in student achievement on CSAP tests in ProComp 2.0 relative to 1.0 years; this is on the order of magnitude of 2 to 3 percent of a standard deviation.

Next we turn our attention to the findings that include teacher fixed effects in the model. These models help to identify whether any ProComp effects may be do to changes in individual teacher productivity. This is an important area of investigation since the adoption of the ProComp system entails more than just changes to teacher pay structure (see Chapter 3 for more detail). In particular, the original ProComp agreement required a revamped performance evaluation system of all DPS teachers, not just those who are enrolled in the pay reform. And, as Taylor and Tyler (2011) find, evaluation may lead to significant improvements in teacher performance by providing them with substantive feedback on how they might improve.

In the teacher fixed effects specification of the model, the magnitudes of the ProComp 2.0 coefficients (columns 2 and 4), while still positive, are greatly reduced and no longer statistically significant. This suggests that the productivity of individual teachers at the elementary level does not increase (i.e. there is no within teacher ProComp effect) as a consequence of the ProComp system. This is perhaps not surprising given that the comparison here is between ProComp 1.0 and 2.0, and the significant change to the evaluation system occurred upon the adoption of the ProComp system, implying that we might only expect to observe within teacher productivity effects (associated with the evaluation system) when comparing pre-ProComp teacher performance to performance under ProComp.

At the middle and high school levels we do have student achievement data for pre-ProComp periods so in the models for students at the secondary level we designate the omitted reference category as the Pre-ProComp years of data (school years 2003-04 and 2004-05). We report the results of these models in Table 3. Columns 1-4 of Panel A show the estimated ProComp coefficients for math achievement at middle school level with teacher covariates (columns 1 and 3) and teacher fixed effects (columns 2 and 4), and columns 5-8 show the analogous results for reading models. Comparable results for the high school level are reported in Panel B of the table.\textsuperscript{35}

We begin by focusing on models that include a single ProComp indicator and teacher covariates (columns 1 and 5). These specifications suggest significant and positive ProComp effects at both the middle and high school levels for both math and reading, and the effects are larger at the high school than middle school level.

\textsuperscript{35} While not reported, these models include the same set of student and teacher covariates as those in Table 2 above.
As described above, there was a significant revision in the ProComp system in 2008. We investigate whether these changes to ProComp are associated with student achievement by substituting separate indicators for ProComp 1.0 and 2.0 for the single ProComp indicator. The results from this specification when we include teacher covariates (columns 3 and 7) indicate heterogeneous math effects across ProComp years. In particular, we find marginally significant (at the 10 percent level) negative effects of ProComp 1.0 relative to pre-ProComp in math at the middle school level, but significantly positive ProComp 2.0 effects. In reading at the middle school level the results are consistently positive and are about the same magnitude. At the high school level, both ProComp periods are significant and positive relative to pre-ProComp years in both math and reading, but the relative effect of ProComp 2.0 is significantly larger than ProComp 1.0 in reading.

As we did at the elementary level, we also estimate teacher fixed effects specifications of our model, where the ProComp coefficients are identified based on within teacher variation in student achievement. These findings are reported in columns 2 and 4 for math and 6 and 8 for reading. While there is some evidence in math, that the ProComp results from the teacher fixed effects specification differ from the specification that includes teacher covariates, no clear pattern emerges for the change in findings. For instance, at the middle school level the ProComp coefficient changes from positive and marginally significant (in column 1) to negative (in column 2), but at the high school level the magnitude of the coefficient increases somewhat when moving from the teacher covariate to fixed effects specification. In reading, a clearer pattern is apparent; the ProComp indicators remain significant and positive at both the middle and high school levels, but the magnitudes are slightly smaller at the high school level.

Under the specification that includes ProComp 1.0 and 2.0 indicators and teacher fixed effects, the math coefficients (in column 4) are smaller than in the covariate model (in column 3) in both ProComp periods in middle schools, but little changed at the high school level. The reading results under the teacher fixed effects specification (column 8) did not change significantly from the teacher covariate findings (column 7). The fact that some of the ProComp coefficients remain statistically significant and positive in specifications that include teacher fixed effects suggests that productivity changes in DPS are not driven solely by changes in teacher workforce composition (i.e. there is evidence of individual teacher productivity changes).

While not reported, we estimate models that include interactions between ProComp variables and student demographics to identify whether there are differential ProComp effects for various student sub-groups. There were no consistent patterns across sub-groups and grades. In most cases the interaction terms are not statistically significant and when they are, their magnitude and direction vary across grade level and subject area.

In summary, looking across Tables 2 and 3, student achievement during ProComp years increased relative to the baseline pre-ProComp years (or in the case of elementary schools, in 2.0 relative to 1.0). The increase is especially large at the secondary level for reading under ProComp 2.0. We wish to be cautious about drawing any strong conclusions about the efficacy

---

36 The differential between ProComp 1.0 and 2.0 is statistically significant at the .05 level in all models except for the teacher fixed effects specification for middle school reading.
of ProComp given that our analysis is based solely on within district change. However, there is some evidence (e.g. Wiley et al., 2010) that DPS has made important progress since the implementation of ProComp. We illustrate this in Figure 4, which shows the differential between the average statewide performance of students on the CSAP tests and performance in several large urban districts in the Denver metro area (Adams 12 Five Star, Aurora Public Schools, and Denver Public Schools).

On the whole this figure reflects our above findings. Across elementary, middle, and high school levels, student performance in DPS gained relative to the rest of the state (though in each case students scored lower than the state average). Moreover, the gains were especially large at the middle and high school levels where DPS closed the gap with the state average faster during the ProComp period than did Adams 12 Five Star or Aurora Public Schools.

B. ProComp Enrollment and the Performance of Voluntary Participants

While the above findings suggest student gains under the ProComp system, they do not specify whether those gains are a function of teachers who are actually enrolled in the alternative pay system or are more reflective of broader changes in DPS. Research suggests that performance management systems in public schools are generally weak (DeArmond et al., 2008). While much of the focus on ProComp is related to its alternative compensation elements, it is also the case that the ancillary systems that are required for ProComp’s operation are quite different from what is necessary to administer pay under the single salary schedule. And it is conceivable that these systems help to improve system operation (e.g. by providing clearer performance feedback to teachers) that improves the productivity of all teachers, not just those participating in the alternative pay system.

We explore this issue by estimating models (consistent with equation 3 above) that include indicators for whether teachers are enrolled in the alternative pay system and if so, whether this enrollment was voluntary or compulsorily.

Recall that at the elementary level we do not have student achievement information for the pre-ProComp years, so the omitted reference category for teacher comparisons at the elementary level is non-participant teachers. Participant and voluntary status indicators represent comparisons amongst teachers who are all in DPS during ProComp years. In the simplest models estimating ProComp participation (columns 1 and 2 in math and 4 and 5 in reading), the ProComp participant variables are insignificant in both subjects whether or not the model includes covariates or teacher fixed effects.

It is important to note that the identification for the ProComp participant coefficient in the fixed effects specifications (columns 2 and 5) comes solely from teachers who opt in. The fact that these coefficients are insignificant suggests that there is not an individual teacher

37 Here the variable “ProComp participant” shows the estimated student achievement for students with a teacher enrolled in ProComp relative to one who is not enrolled, and the “Voluntary participant” variable the estimated student achievement for students with a teacher voluntarily in ProComp relative to one who is compulsorily enrolled in the system. The estimated effect of having a teacher who is voluntarily enrolled in ProComp relative to one who is not enrolled is the sum of “ProComp participant” and “Voluntary participant.”
productivity impact associated with ProComp participation in math or reading at the elementary level.

We expect effective teachers who might gain from a pay for performance system to be more likely to opt in to ProComp, and there is some evidence that voluntary ProComp participants differ from teachers who are hired into the ProComp system in terms of attitudes and instructional behavior (Wiley et al., 2010). Given this it would not be surprising to find systematic differences in teacher productivity based on opt in status. The positive and nearly significant (p=.105) coefficient on the voluntary participant indicator at the elementary level in math supports this hypothesis and suggests that teachers who voluntarily opted in were more effective that non-voluntary teachers before they opted in. None of the participant or voluntary status results are significant in reading at the elementary level.

Table 5 reports results the middle (Panel A) and high school results (Panel B) for math (columns 1–3) and reading (columns 4–6). The reference category for these models is teachers in pre-ProComp years. In contrast to the results at the elementary level, here we do see evidence of differential productivity effects of ProComp teachers (again, note here that the reference group is different than for elementary teachers), but the results are puzzling in that no consistent pattern emerges across grade configuration and subject area.

Focusing first at the middle school level in math, there is strong evidence in both teacher covariate (column 1) and fixed effects (column 2) model specifications that the overall DPS gains in student achievement reported in the prior sub-section are driven by teachers who are participating in the alternative pay system. Specifically, the achievement of students assigned to non-ProComp teachers in years in which ProComp is in effect is estimated to be slightly lower (about 2 to 3 percent of a standard deviation) than the achievement prior to the implementation of the ProComp system. The achievement of students assigned to a teacher participating in the ProComp system, however, is significantly higher (by 6 to 7 percent of a standard deviation) than that of non-participating teachers, so much so that the net effect of having a ProComp participating teacher is estimated to be positive relative to years prior to the implementation of ProComp.

The net effect achievement level in reading for students in ProComp years is also positive, but here the ProComp effect is driven by non-participant teachers. Note, for instance, the significant positive coefficient for “ProComp years” (in both covariate and fixed effects

---

38 We find further support for this conclusion with models not reported here that compares the effectiveness of teachers who chose to opt in with the effectiveness of teachers who had the choice to opt in and chose not to.

39 Here the “ProComp years” variable indicates the estimated achievement of students with non-ProComp teachers who are teaching in ProComp system years (relative to teachers in pre-ProComp years), and the participant and voluntary variables are still interpreted in the same way in the elementary level models described above. In these models the net effect of having a ProComp participating teacher, relative to student achievement in pre-ProComp years, is the sum of “ProComp years” and “ProComp participant,” and the net effect of having a teacher who voluntarily participates in the system is the sum of these two coefficients plus “Voluntary participant.”

40 The effect of assignment to a non-ProComp teacher during the period when ProComp is in effect is identified by the “ProComp years” variable and the effect of assignment to a ProComp teacher during the ProComp period is the sum of the “ProComp years” and “ProComp participant” coefficients.
specifications) and negative coefficient on “ProComp participant” (in both specifications and significant in the fixed effects specification).

At the high school level the pattern changes. The net effect of having either a ProComp participating teachers or non-participating teacher in a year in which ProComp is in effect in DPS is positive in both math and reading.\(^{41}\) In math, the achievement of students who have ProComp participating teachers is slightly lower than the achievement of students with non-participating teachers, whereas in reading the achievement of students with ProComp participating teachers is significantly higher than the achievement of students with non-participating teachers. Again, this pattern of results holds for both specifications with teacher covariates and specifications with teacher fixed effects.

Finally, we turn our attention to the specifications that include indicators for a teacher’s voluntary status in ProComp (column 3 for math and 6 for reading). These results show no clear pattern. In middle school math, ProComp participants are more effective than non-participants, with compulsory ProComp teachers performing slightly better than voluntary ProComp teachers. In middle school reading, non-ProComp teachers are more effective and there is no statistically significant difference between voluntary and compulsory ProComp teachers. At the high school level, compulsory ProComp teachers are less effective in math than both voluntary ProComp and non-ProComp teachers. In reading, compulsory and voluntary ProComp teachers perform better than non-ProComp teachers.\(^{42}\)

Wiley et al. (2010) also investigates the effectiveness of teachers voluntarily opting into ProComp and finds that voluntary ProComp teachers are more effective than both compulsory ProComp participants and non-ProComp teachers.\(^{43}\) The findings that we described above are obviously less straightforward. There are at least two potential explanations for what appears to be a divergence in results. First, our analysis includes the 2009-10 school year that was not part of Wiley et al.’s sample. Second, it is possible that the effects of voluntarily opting in are heterogeneous across grade levels; Wiley et al.’s analysis does not allow for differentiated effects across grade configurations while ours does. We check for the source of divergence by estimating specifications that do not differentiate across grade levels and that use a sample consistent with Wiley et al.

We largely replicate Wiley et al.’s findings when we estimate models that do not allow for differentiated effects by grade level and exclude the 2009-10 school year from the analysis.\(^{44}\) When we include all years and all grades together, we find that in math, compulsory ProComp teachers are more effective than non-ProComp participants, and voluntary teachers are more effective than both ProComp and non-ProComp teachers. ProComp participants and non-ProComp teachers show no statistically significant difference in reading with this specification. These results suggest that the divergence between our findings and those of Wiley et al. are

---

\(^{41}\) This carries forward from the results in the above sub-section, but is also seen by adding the coefficients on “ProComp years” to “ProComp participant.”

\(^{42}\) We also estimated these same sets of models for various student sub-groups. As we described in sub-section A, there was no clear pattern of findings for these subgroups.

\(^{43}\) His outcome measure in this analysis is the median Colorado Growth Percentile score for each teacher.

\(^{44}\) We get a similar pattern of results, although with slightly different ProComp effect sizes, which is not surprising given that Wiley et al. do not include teacher covariates in their model.
explained by the additional 2009-10 school year, which in turn implies that teachers hired in 2009-10 (who are compulsorily enrolled in ProComp) are relatively effective. When we exclude the 2009-10 school year from our analytic sample and estimate the same regressions separately by grade configuration, we find continue to find heterogeneous effects across grade levels, which draws attention to the importance of allowing for heterogeneous ProComp effects across grade levels.

**C. How Effectively Are ProComp Rewards Targeted?**

We begin the exploration of the targeting of awards by ranking all teachers based on the teacher effectiveness estimates generated by Equation (4) then focusing on the set of teachers who receive each of the individual teacher ProComp incentives to determine into which effectiveness quintile they fall. Table 6 shows these results by incentive type and year. To the degree that the value-added rankings correspond to the receipt of a ProComp incentive, we would expect to see far more teachers who receive incentives in the upper quintiles than lower quintiles.

As would be expected given the correlation between the EE incentive and value-added estimates (Panel A), teachers receiving this award are disproportionately in the upper end of the value-added distribution; for example, in math, less than 25 percent of teachers receiving this incentive are in the lowest two quintiles in each year and over 60 percent are in the top two quintiles. The distribution is less skewed for reading, but is still weighted toward the top.

The pattern of results is less clear for the SGO incentives (Panel B), though in general, we observe a slightly higher proportion of teachers at the top of the distribution than the bottom. For instance, across all years the total in the top two quintiles is 42 percent versus 37 percent in the bottom two quintiles (the corresponding figures for reading are 41 and 38).

For PDUs and CPEs (Panels C and D), there does not appear to be much evidence at all that teachers receiving these incentives are more effective in either math or reading. For instance, in some years the teachers receiving these incentives tend to be at the top of the value-added distribution, but in other years they are more likely to be at the bottom.

We investigate these findings further by estimating logistic regressions predicting the likelihood of award receipt as a function of teacher effectiveness. Table 7 reports the marginal effect estimates for these models. Columns 1 and 2 show the results for teacher effectiveness measured based on student achievement in math, and columns 3 and 4, for teacher effectiveness measured based on student achievement in reading.

---

45 Note that the timing of receipt of a bonus corresponds to the school year in which the bonus was earned, not when the year in which it was paid out.

46 We would expect recipients to be scattered evenly across quintiles if the receipt of incentives were random.

47 The results in Table 7 are robust to models that include teacher covariates (experience, experience-squared, and degree). When these models are run separately by grade configuration, the magnitudes and significance levels change somewhat, but the overall patterns remain.
Several findings are noteworthy. There is strong evidence that EE and SGO awards are associated with teacher effectiveness (i.e. value-added) in both math and reading. This is true whether effectiveness is based on comparisons within and across schools (columns 1 and 3) or only within schools (columns 2 and 4) as is the case when the models include school fixed effects. 48

Given the high correlations between EE and value-added, it is not surprising to find the receipt of the EE award to be associated with teacher effectiveness. There is, however, little existing research on the efficacy of SGO-type award systems so the finding on SGOs has potentially important policy implications for states and localities wishing to reward teachers in grades or subjects not covered by state assessments (we discuss this further below).49

There is relatively little evidence, however, that the CPE and PDU ProComp awards are related to teacher effectiveness, though a case can be made for CPEs in math, where the coefficient on teacher effectiveness is quite small, but statistically significant.50

The above findings are illustrated in Figure 5, which shows kernel distributions of value-added (in math and reading) for each incentive. The dashed lines represent the distributions for those teachers who receive an award and the solid line ProComp teachers who do not, with drop down lines to indicate the mean of each distribution.

The separation between the dotted and solid distributions for the EE and SGO awards is clear (i.e. significant weight of the dashed distribution lies to the right of the solid distribution). But, in the case of CPEs and PDUs, there is almost complete overlap of the distributions of effectiveness for those who do and do not receive these awards. We began this sub-section by noting that a one standard deviation change in teacher effectiveness translates into roughly 0.15 standard deviations of student achievement. This provides a means of gauging the average differential between award recipients and non-recipients. Specifically, the differential for the EE award is about a full standard deviation in terms of math achievement and nearly 60 percent of a standard deviation for reading achievement. The SGO differentials are smaller, but still about 35 percent of a standard deviation for math and 20 percent for reading. The differentials for CPEs and PDUs are negligible and, again, only statistically significant in the case of the math CPE.

Given that CPE and PDU awards are not strongly associated with teacher effectiveness, it is worth digging a bit deeper into the results. In particular, the CPE results are intriguing because there is some existing evidence in the literature that principals can distinguish teacher effectiveness (Jacob and Lefgren, 2008, Jacob, 2011), even if differences in effectiveness are

---

48 We also estimate models that predict the probability of receiving an SGO bonus based on effectiveness quintiles rather than the linear effectiveness estimate and find similar results.
49 The only other evidence on SGOs that we are aware of is from Austin, TX, which uses SGOs as part of the REACH program. Research by the Austin school district suggests that teachers whose students met learning objectives were more likely to raise student achievement on the state standardized test (Schmitt and Ibanez, 2011).
50 We also run regressions that include interactions for effectiveness and probationary status but find no evidence of an interaction effect. This suggests that the effect of teacher effectiveness on the probability of receipt of each bonus is similar for probationary and non-probationary teachers.
rarely documented (Weisburg et al., 2009). And, in the case of PDUs, we might not expect much of a relationship to effectiveness since these awards are intended to provide teachers with a financial incentive to build their human capital through training (see Chapter 4). Thus, if successful, teachers might be more effective after having completed PDUs.

As noted above, we do find the probability of receiving a CPE award rises marginally with teacher effectiveness in math. One possibility is that the linear specification masks stronger relationships at points along the teacher performance distribution. We test this by including indicators for the effectiveness quintile in which each teacher falls (with the bottom quintile being the reference group). In this specification, teachers falling into quintiles other than the bottom have a higher probability of reward receipt, with a marked increase in that probability for teachers in the top quintile. None of the quintiles for reading effectiveness are statistically significant.

We test the human capital hypothesis for PDUs by regressing student achievement on the number of PDUs a teacher earned in the previous year (and whether a teacher has a masters degree). Regardless of how these models are specified (e.g. with or without school fixed effects, whether the effect of increased PDUs is assumed to be linear, etc.) PDUs are not significant predictors of student achievement in either math or reading. Again, it may be that this award is designed to develop teachers in ways that are not easily detected through analysis of student test achievement, but it is clear from all the findings on PDUs that this is not an award that shows up as significant in predicting students’ CSAP scores.

Finally, while not all awards are related to estimated effectiveness, it is conceivable that the targeting of awards improved over time as DPS developed more sophisticated means of assessing teachers. This is not born out in the data. Specifically, as we report in the table, few of the effectiveness year interaction terms are statistically significant, and those that are significant are negative. If anything, the relationship between effectiveness and the receipt of the awards tended to be weaker in the more recent ProComp years than when the ProComp system was first implemented, which also corresponds to a period that showed a large increase in the number of ProComp awards teachers received.

V. Discussion and Conclusions

Our findings document significant student learning gains in DPS across grades and subjects. The source of those gains, however, are not altogether clear as there is not a consistent

---

51 It is of course entirely possible that any of the ProComp awards, other than EE reward aspects of teaching quality that are not strongly associated with student test achievement (so would not necessarily show up in value-added estimates).

52 The Jacob and Leftgren study, for instance, suggests that principals are better able to distinguish teachers at the top and bottom of the performance distribution than the middle (Jacob and Leftgren, 2008)

53 The marginal effect estimates for quintiles 2–5 are: .16, .50, .34, and .83 for the models where teacher effectiveness is based on the model without school fixed effects, with quintiles 3, 4, and 5 being statistically significant. In the second model based on within-school teacher effectiveness the marginal effect estimates are .27, .41, .52, and .88 (quintiles 3, 4, and 5 are significant).

54 This is also true of Masters degrees, again, a common finding in the literature (Goldhaber and Brewer, 1997; Harris and Sass, 2011).
pattern across grade level and subject: in some cases the gains appear primarily amongst students with ProComp teachers, while in other cases non-ProComp teachers are found to be more effective. Though puzzling, these findings are not inconsistent with research on other well-known interventions that include elements similar to ProComp.

Taylor and Tyler’s (2011) investigation of the effects of evaluation reform show large student achievement effects in math, but no effect on students’ reading performance. And, research on another well-known pay reform system – the ‘Teacher Advancement Program (TAP)’ – finds that the student achievement effects of are heterogeneous across grade levels, with positive and significant effects associated with being a TAP school at the elementary level, and mixed and sometimes significantly negative effects at the secondary level (Springer et al., 2010).55 The authors postulate that incentives at the secondary level may lead to a reduction in teamwork. It is not inconceivable that this could also be true for DPS, but the way the incentives affects teacher teams would have to be more complex in that we found positive effects in some cases for teachers who were not enrolled in ProComp, implying that any negative spillovers did not severely impact their performance. And, we found heterogeneous effects across subjects. While it is simply further speculation, an additional possibility is that ProComp could be unevenly implemented across grade levels and subjects. Unfortunately our findings on this are limited due to a lack of information on the quality of ProComp implementation across schools, moreover our findings are possibly confounded by other reform efforts that tend to be targeted to particular grades and subjects.56

Two other results that have potentially far-reaching policy implications are clearer. The first is that “ProComp effects” are not always concentrated solely amongst teachers enrolled in ProComp (this is especially true at the high school level). This suggests that systems associated with ProComp implementation, e.g. data and evaluation, may have beneficial spillovers to non-ProComp teachers. The second is that some of the ProComp awards do successfully target teacher effectiveness. The finding that SGOs appear to be a reasonably successful means of rewarding teachers whose students demonstrate larger than expected gains on the CSAP tests provides support for other states and localities (e.g., Race to the Top) looking to reform teacher incentives. However, several of the ProComp bonuses appear unrelated to either current or future teacher value-added measurements. Whether this is good or bad is clearly a normative question as some might argue that these awards are rewarding aspects of classroom instruction not strongly associated with students’ test achievement.

References

Ballou, Dale, William L. Sanders, and Paul S. Wright. "Controlling for Student Background in Value-

---

55 TAP is not directly comparable to ProComp, but like ProComp it is a comprehensive reform that entails changes to teacher evaluation and advancement.
56 Aside from collecting additional information on these possibilities, one avenue of potentially fruitful future research would be to assess whether the concentration of ProComp teachers in schools influences the findings as this may speak to the degree to which the incentive influences teacher teams.


Center on Performance Incentives, 2009.
Figures and Tables
Figure 1. Percent of ProComp Teachers Receiving an Award by Category

Source: DPS Payments data
Note: Not all awards were available in all years. Specifically, awards for professional development units, student growth objectives, exceeds CSAP expectations and high growth schools were not available in 2005-06. Awards for high growth schools were also not available in 2006-07.

Figure 2. Percent of Teachers in ProComp by Voluntary Status

Source: DPS Payments data
Figure 3. Causal Pathways Through which ProComp May Affect Teacher Workforce Quality

A. Improve the pool of newly trained teachers
B. Improve the effectiveness of newly hired teachers
C. Improve the effectiveness of existing teachers
D. Retain the most effective teachers
E. Exit Less Effective Teachers
Figure 4. Average Difference between State and Selected District CSAP Scores over Time

Panel A: Elementary, Reading

<table>
<thead>
<tr>
<th></th>
<th>Pre-ProComp</th>
<th>ProComp 1.0</th>
<th>ProComp 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: CSAP scores are vertically scaled. Data from Colorado Department of Education, www.schoolview.org, Data Lab data reporting system. This system includes a disclaimer, “because of the number and types of data aggregations available here, many results have not been verified.”
Figure 5. Kernel Density for Teacher Effectiveness by Award Status

Kernel Density Distribution of Teacher Effectiveness in Math for Exceeds Expectations Bonus

Mean Differential = 0.130

Kernel Density Distribution of Teacher Effectiveness in Reading for Exceeds Expectations Bonus

Mean Differential = 0.085
Kernel Density Distribution of Teacher Effectiveness in Math for Student Growth Objective Bonus

Mean Differential = 0.055

Kernel Density Distribution of Teacher Effectiveness in Reading for Student Growth Objective Bonus

Mean Differential = 0.033
Kernel Density Distribution of Teacher Effectiveness in Math for Professional Development Unit Bonus

Kernel Density Distribution of Teacher Effectiveness in Reading for Professional Development Unit Bonus

Mean Differential = 0.014
Table 1. Student Sample Statistics for Unrestricted and Restricted Samples

<table>
<thead>
<tr>
<th></th>
<th>Unrestricted Sample</th>
<th></th>
<th></th>
<th>Restricted Sample</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Native American</td>
<td>0.012</td>
<td>0.107</td>
<td>0.012</td>
<td>0.110</td>
<td>0.012</td>
<td>0.108</td>
</tr>
<tr>
<td>Asian</td>
<td>0.054</td>
<td>0.227</td>
<td>0.057</td>
<td>0.231</td>
<td>0.037</td>
<td>0.188</td>
</tr>
<tr>
<td>Black</td>
<td>0.159</td>
<td>0.365</td>
<td>0.156</td>
<td>0.363</td>
<td>0.178</td>
<td>0.382</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.559</td>
<td>0.496</td>
<td>0.558</td>
<td>0.497</td>
<td>0.580</td>
<td>0.494</td>
</tr>
<tr>
<td>White</td>
<td>0.196</td>
<td>0.397</td>
<td>0.182</td>
<td>0.386</td>
<td>0.194</td>
<td>0.395</td>
</tr>
<tr>
<td>Free/reduced price lunch</td>
<td>0.661</td>
<td>0.473</td>
<td>0.659</td>
<td>0.474</td>
<td>0.681</td>
<td>0.466</td>
</tr>
<tr>
<td>Learning disability</td>
<td>0.122</td>
<td>0.327</td>
<td>0.116</td>
<td>0.320</td>
<td>0.136</td>
<td>0.343</td>
</tr>
<tr>
<td>Standardized math CSAP score</td>
<td>0.000</td>
<td>1.000</td>
<td>0.071</td>
<td>0.982</td>
<td>-0.023</td>
<td>0.998</td>
</tr>
<tr>
<td>Standardized reading CSAP score</td>
<td>0.000</td>
<td>1.000</td>
<td>-0.021</td>
<td>0.979</td>
<td>-0.059</td>
<td>1.003</td>
</tr>
<tr>
<td>Observations</td>
<td>183,549</td>
<td></td>
<td>215,971</td>
<td></td>
<td>155,256</td>
<td></td>
</tr>
</tbody>
</table>

Note: the unrestricted sample includes student-year observations from 2004-2010 in grades 4-10 for math and reading.

The means for the restricted sample are based on student-year-subject observations. Students may be counted multiple times in the restricted sample because they can have both ProComp and non-ProComp teachers in the same year.
### Table 2. Estimated Student Achievement under ProComp: Elementary School Level

<table>
<thead>
<tr>
<th>ProComp Variables</th>
<th>Math</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>ProComp 2.0 (SY2007-08 to 2009-10)</td>
<td>0.0202***</td>
<td>0.00434</td>
</tr>
<tr>
<td></td>
<td>(0.00509)</td>
<td>(0.00619)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student Variables</th>
<th>Math</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior-year math score</td>
<td>0.736***</td>
<td>0.732***</td>
</tr>
<tr>
<td></td>
<td>(0.00386)</td>
<td>(0.00379)</td>
</tr>
<tr>
<td>Prior-year reading score</td>
<td>0.116***</td>
<td>0.110***</td>
</tr>
<tr>
<td></td>
<td>(0.00410)</td>
<td>(0.00398)</td>
</tr>
<tr>
<td>Native American</td>
<td>-0.0899***</td>
<td>-0.0549**</td>
</tr>
<tr>
<td></td>
<td>(0.0253)</td>
<td>(0.0242)</td>
</tr>
<tr>
<td>Asian</td>
<td>0.0838***</td>
<td>0.0946***</td>
</tr>
<tr>
<td></td>
<td>(0.0150)</td>
<td>(0.0145)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.157***</td>
<td>-0.116***</td>
</tr>
<tr>
<td></td>
<td>(0.00979)</td>
<td>(0.0102)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.0931***</td>
<td>-0.0422***</td>
</tr>
<tr>
<td></td>
<td>(0.00826)</td>
<td>(0.00866)</td>
</tr>
<tr>
<td>Free/reduced price lunch</td>
<td>-0.0498***</td>
<td>-0.0327***</td>
</tr>
<tr>
<td></td>
<td>(0.00710)</td>
<td>(0.00722)</td>
</tr>
<tr>
<td>Learning disability</td>
<td>-0.188***</td>
<td>-0.183***</td>
</tr>
<tr>
<td></td>
<td>(0.00812)</td>
<td>(0.00789)</td>
</tr>
<tr>
<td>5th Grade</td>
<td>-0.0135**</td>
<td>-0.000374</td>
</tr>
<tr>
<td></td>
<td>(0.00525)</td>
<td>(0.00937)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Teacher Variables</th>
<th>Math</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Year Prior Experience</td>
<td>0.0368***</td>
<td>Fixed Effect</td>
</tr>
<tr>
<td></td>
<td>(0.0111)</td>
<td></td>
</tr>
<tr>
<td>2 Years Prior Experience</td>
<td>0.0449***</td>
<td>0.0360***</td>
</tr>
<tr>
<td></td>
<td>(0.0118)</td>
<td></td>
</tr>
<tr>
<td>3 Years or More Prior Experience</td>
<td>0.0607***</td>
<td>0.0406***</td>
</tr>
<tr>
<td></td>
<td>(0.00840)</td>
<td></td>
</tr>
<tr>
<td>Masters or higher</td>
<td>0.000262</td>
<td>-0.00985*</td>
</tr>
<tr>
<td></td>
<td>(0.00521)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>36,710</td>
<td>36,710</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.762</td>
<td>0.796</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
All specifications include grade level and missing value dummy variables.
<table>
<thead>
<tr>
<th>Table 3. Estimated Student Achievement under ProComp: Middle and High School Level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Middle Schools</strong></td>
</tr>
<tr>
<td><strong>Math</strong></td>
</tr>
<tr>
<td>ProComp years (SY2005-06 to 2009-10)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ProComp 1.0 (SY2005-06 and 2006-07)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ProComp 2.0 (SY2007-08 to 2009-10)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Teacher Controls</td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>R-squared</td>
</tr>
<tr>
<td><strong>Panel B: High Schools</strong></td>
</tr>
<tr>
<td><strong>Math</strong></td>
</tr>
<tr>
<td>ProComp Years (SY2005-06 to 2009-10)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ProComp 1.0 (SY2005-06 and 2006-07)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ProComp 2.0 (SY2007-08 to 2009-10)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Teacher Controls</td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>R-squared</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications also include prior-year test scores, free/reduced price lunch status, IEP status, race/ethnicity, grade level, and missing value dummy variables. There are 454 teachers who are in the analytic sample in both pre-ProComp years and ProComp years. In columns 3, 4, and 7, the difference between the ProComp 1.0 effect and the ProComp 2.0 effect is statistically significant at the .05 level; in column 8 the difference is not statistically significant.
Table 4. Student Achievement by ProComp Status: Elementary School Level

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProComp 2.0 (SY2007-08 to 2009-10)</td>
<td>0.0201***</td>
<td>0.00629</td>
<td>0.0211***</td>
<td>0.0295***</td>
<td>0.00648</td>
<td>0.0297***</td>
</tr>
<tr>
<td></td>
<td>(0.00521)</td>
<td>(0.00606)</td>
<td>(0.00525)</td>
<td>(0.00563)</td>
<td>(0.00734)</td>
<td>(0.00568)</td>
</tr>
<tr>
<td>Procomp participant</td>
<td>0.000819</td>
<td>-0.0148</td>
<td>-0.00811</td>
<td>0.00448</td>
<td>0.00324</td>
<td>0.00332</td>
</tr>
<tr>
<td></td>
<td>(0.00545)</td>
<td>(0.0175)</td>
<td>(0.00775)</td>
<td>(0.00588)</td>
<td>(0.0200)</td>
<td>(0.00834)</td>
</tr>
<tr>
<td>Voluntary</td>
<td>0.0143</td>
<td>0.00186</td>
<td>0.00881</td>
<td>0.00951</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>36,710</td>
<td>36,710</td>
<td>36,710</td>
<td>33,129</td>
<td>33,129</td>
<td>33,129</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.762</td>
<td>0.796</td>
<td>0.762</td>
<td>0.723</td>
<td>0.745</td>
<td>0.723</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications also include prior-year test scores, free/reduced price lunch status, IEP status, race/ethnicity, grade level, and missing value dummy variables. The omitted reference category is ProComp1.0 time period.

Table 5. Student Achievement by ProComp Status: Middle and High School Level

Table 5. Student Achievement by ProComp Status: Middle and High School Level

Panel A: Middle Schools

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProComp years (SY2005-06 to 2009-10)</td>
<td>-0.0174***</td>
<td>-0.0361***</td>
<td>-0.0179***</td>
<td>0.0317***</td>
<td>0.0353***</td>
<td>0.0317***</td>
</tr>
<tr>
<td></td>
<td>(0.00620)</td>
<td>(0.00829)</td>
<td>(0.00620)</td>
<td>(0.00608)</td>
<td>(0.00771)</td>
<td>(0.00608)</td>
</tr>
<tr>
<td>Procomp participant</td>
<td>0.0628***</td>
<td>0.0666***</td>
<td>0.0803***</td>
<td>-0.00334</td>
<td>-0.0399***</td>
<td>-0.000598</td>
</tr>
<tr>
<td></td>
<td>(0.00486)</td>
<td>(0.01000)</td>
<td>(0.00699)</td>
<td>(0.00438)</td>
<td>(0.0111)</td>
<td>(0.00612)</td>
</tr>
<tr>
<td>Voluntary</td>
<td>-0.0268***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00773)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>55,626</td>
<td>55,626</td>
<td>55,626</td>
<td>69,543</td>
<td>69,543</td>
<td>69,543</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.769</td>
<td>0.791</td>
<td>0.769</td>
<td>0.733</td>
<td>0.750</td>
<td>0.733</td>
</tr>
</tbody>
</table>

Panel B: High Schools

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProComp years (SY2005-06 to 2009-10)</td>
<td>0.0796***</td>
<td>0.0938***</td>
<td>0.0790***</td>
<td>0.0504***</td>
<td>0.0338***</td>
<td>0.0499***</td>
</tr>
<tr>
<td></td>
<td>(0.00751)</td>
<td>(0.00945)</td>
<td>(0.00751)</td>
<td>(0.00697)</td>
<td>(0.00829)</td>
<td>(0.00697)</td>
</tr>
<tr>
<td>Procomp participant</td>
<td>-0.0137**</td>
<td>-0.0243*</td>
<td>-0.0274***</td>
<td>0.0245***</td>
<td>0.0520***</td>
<td>0.0391***</td>
</tr>
<tr>
<td></td>
<td>(0.00647)</td>
<td>(0.0126)</td>
<td>(0.00865)</td>
<td>(0.00716)</td>
<td>(0.0163)</td>
<td>(0.0103)</td>
</tr>
<tr>
<td>Voluntary</td>
<td>0.0244**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0102)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>32,315</td>
<td>32,315</td>
<td>32,315</td>
<td>37,098</td>
<td>37,098</td>
<td>37,098</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.737</td>
<td>0.749</td>
<td>0.738</td>
<td>0.695</td>
<td>0.709</td>
<td>0.695</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications also include prior-year test scores, free/reduced price lunch status, IEP status, race/ethnicity, grade level, and missing value dummy variables. The omitted reference category is the pre-ProComp time period.
Table 6. Percent Receiving Bonus by Quintile of Effectiveness

<table>
<thead>
<tr>
<th>Panel A. Exceeds Expectations</th>
<th>Math</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quintiles</td>
<td>All</td>
<td>2007</td>
</tr>
<tr>
<td>lowest - 1</td>
<td>6.8</td>
<td>2.0</td>
</tr>
<tr>
<td>2</td>
<td>12.5</td>
<td>8.2</td>
</tr>
<tr>
<td>3</td>
<td>18.5</td>
<td>22.5</td>
</tr>
<tr>
<td>4</td>
<td>24.1</td>
<td>16.3</td>
</tr>
<tr>
<td>highest - 5</td>
<td>38.1</td>
<td>51.0</td>
</tr>
<tr>
<td>Total teachers</td>
<td>680</td>
<td>49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. SGOs</th>
<th>Math</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quintiles</td>
<td>All</td>
<td>2007</td>
</tr>
<tr>
<td>lowest - 1</td>
<td>17.7</td>
<td>19.2</td>
</tr>
<tr>
<td>2</td>
<td>18.9</td>
<td>16.4</td>
</tr>
<tr>
<td>3</td>
<td>19.5</td>
<td>20.6</td>
</tr>
<tr>
<td>4</td>
<td>21.0</td>
<td>21.2</td>
</tr>
<tr>
<td>highest - 5</td>
<td>23.0</td>
<td>22.6</td>
</tr>
<tr>
<td>Total teachers</td>
<td>1326</td>
<td>146</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C. PDUs</th>
<th>Math</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quintiles</td>
<td>All</td>
<td>2007</td>
</tr>
<tr>
<td>lowest - 1</td>
<td>18.2</td>
<td>22.3</td>
</tr>
<tr>
<td>2</td>
<td>20.1</td>
<td>17.0</td>
</tr>
<tr>
<td>3</td>
<td>19.6</td>
<td>17.0</td>
</tr>
<tr>
<td>4</td>
<td>20.8</td>
<td>23.4</td>
</tr>
<tr>
<td>highest - 5</td>
<td>21.3</td>
<td>20.2</td>
</tr>
<tr>
<td>Total teachers</td>
<td>942</td>
<td>94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel D. CPEs</th>
<th>Math</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quintiles</td>
<td>All</td>
<td>2007</td>
</tr>
<tr>
<td>lowest - 1</td>
<td>19.2</td>
<td>21.1</td>
</tr>
<tr>
<td>2</td>
<td>19.5</td>
<td>16.9</td>
</tr>
<tr>
<td>3</td>
<td>20.6</td>
<td>20.4</td>
</tr>
<tr>
<td>4</td>
<td>19.4</td>
<td>21.1</td>
</tr>
<tr>
<td>highest - 5</td>
<td>21.3</td>
<td>20.4</td>
</tr>
<tr>
<td>Total teachers</td>
<td>1168</td>
<td>142</td>
</tr>
</tbody>
</table>