In the following report, Hanover Research reviews recent literature on the different types of value-added models in teacher pay-for-performance programs. Hanover Research discusses research on the validity and reliability of value-added models, and on the impact of these programs on student achievement. In the course of this review, Hanover Research identifies best practices for the design and implementation of pay-for-performance programs.
# Table of Contents

Executive Summary and Key Findings ......................................................... 3

INTRODUCTION ..................................................................................... 3

KEY FINDINGS ..................................................................................... 4

Section I: Overview of Pay-for-Performance and Value-Added Models ................................ 5

PAY-FOR-PERFORMANCE SYSTEMS ..................................................................... 5

Varieties of Pay-for-Performance Models ................................................................ 5

Evaluation Models: Validity, Reliability, and Impact ............................................ 6

VALUE-ADDED MODELS ........................................................................... 8

Validity and Reliability of VAMs ....................................................................... 8

Section II: Impacts of Pay-for-Performance and Value-Added Models .................. 10

EFFECTIVENESS OF PAY-FOR-PERFORMANCE PROGRAMS ......................... 10

RESEARCH ON VALUE-ADDED MODELS: EFFECTIVENESS AND BEST PRACTICES ................................................................. 11

Teachers’ Program Understanding .................................................................... 13

Prior Context and School Characteristics .......................................................... 14

Section III: District Profiles .......................................................................... 15

HARRISON SCHOOL DISTRICT TWO, COLORADO .................................... 16

Program Outline ....................................................................................... 16

Program Budget ........................................................................................ 17

Reception and Impact ................................................................................... 18

Lessons Learned .......................................................................................... 18

ROUND ROCK INDEPENDENT SCHOOL DISTRICT, TEXAS ................. 19

Program Outline ....................................................................................... 19

Addressing Communication and Teacher Understanding Issues ......................... 20

Appendix A: Teacher Incentives and Loss Aversion .......................................... 21

Appendix B: Key Details of Harrison County School District’s Pay-for-Performance Plan .. 22

TEACHER PLACEMENT & MOBILITY ....................................................... 22

EVALUATIONS .......................................................................................... 23
EXECUTIVE SUMMARY AND KEY FINDINGS

INTRODUCTION

School districts throughout the country are implementing pay-for-performance models, which seek to reward teachers who are effective in promoting student learning. While these initiatives offer the prospect of incentivizing teachers and improving student achievement, there is little agreement on how best to structure teacher evaluation and reward systems. One popular approach to evaluating teacher effectiveness is to use value-added models (VAMs), which aim to quantify teachers’ contributions to improvements in students’ test scores. This information about teachers’ impact on student achievement can be used to ensure that effective teachers receive higher compensation. Like pay-for-performance models more generally, VAMs have proved controversial.

In this report, Hanover Research reviews the latest research on the use of value added models in pay-for-performance teacher incentive systems. We focus on two aspects of the literature in particular: evidence regarding the effectiveness of VAM-driven pay-for-performance systems in promoting effective teaching, and research on best practices in designing and implementing pay-for-performance programs. Accordingly, this report comprises the following three sections:

- **Section I: Overview of Pay-for-Performance and Value-Added Models** provides an introduction to the nature of pay-for-performance and value-added models, and discusses differences between different versions of each. Hanover Research reviews the major drawbacks that experts have identified regarding pay-for-performance programs and VAMs in particular and describes strategies that program administrators can use to mitigate these disadvantages.

- **Section II: Impacts of Pay-for-Performance and Value-Added Models** reviews the research literature on teacher financial incentive programs, with an emphasis on the most recent evidence. Hanover Research begins with a discussion of research on pay-for-performance programs of all types, followed by a summary of the latest findings regarding VAM-based incentive programs.

- **Section III: Profiles** provides an in-depth look at two districts that have effectively implemented pay-for-performance models: Harrison School District Two, CO; and Round Rock Independent School District, TX. These profiles provide practical examples of best practices in designing, launching, and sustaining pay-for-performance compensation systems using VAMs.
**Key Findings**

- **Research on the effects of pay-for-performance systems and value-added models has yielded mixed and inconclusive results.** Some studies have found improvements in teacher practices and student performance because of pay-for-performance incentives, but other studies find no such result. In particular, several recent, rigorous studies of pay-for-performance programs using value-added models have failed to identify any significant effects on teachers or students. At present, it is unknown whether pay-for-performance programs increase student achievement.

- **Estimates of teaching effectiveness that combine multiple evaluation measures are superior to those that rely on test scores alone.** If performance evaluations are to determine compensation, then performance evaluations should accurately reflect effectiveness. Research suggests that teacher performance ratings are most accurate when they combine value-added estimates with other measures, such as observation scores. School districts have developed and implemented a number of different models for combining evaluation measures into a single effectiveness score, but there is no consensus regarding the optimal weighting of different measures.

- **VAMs can accurately reflect teachers’ impact on student achievement much of the time, but in certain contexts they can produce highly inaccurate estimations of teachers’ effectiveness.** For this reason, many experts caution against using VAMs as the sole basis for high-stakes personnel decisions.

- **Determining the appropriate VAM formula for a given context can require significant technical expertise.** Before implementing a VAM-based teacher evaluation scheme, districts should ensure that staff with the relevant expertise is available not only for development of the model, but also for regular monitoring.

- **District-level plans to implement VAM-based pay-for-performance programs should include strategies for clear and effective communication with teachers about program features.** Communication difficulties, especially about the complex procedures used to estimate value-added or allocate rewards, have plagued several large-scale pay-for-performance programs. Establishing clearly-defined procedures for disseminating program information to teachers and ensuring that teachers have ready access to knowledgeable sources can help ensure that teachers know how they can take full advantage of incentive programs.
SECTION I: OVERVIEW OF PAY-FOR-PERFORMANCE AND VALUE-ADDED MODELS

In this section, Hanover Research provides an overview of types of pay-for-performance teacher incentive systems and Value-Added Models (VAMs).

PAY-FOR-PERFORMANCE SYSTEMS

Pay-for-performance incentive programs provide financial rewards to teachers that exhibit superior performance, usually as measured by student achievement, classroom observation scores, or some combination of the two. Proponents of teacher pay-for-performance systems argue that such systems will improve the effectiveness of the teaching profession in several ways:

- By providing financial incentives for teachers to focus or increase their effort
- By encouraging the development of stronger teaching skills
- By increasing incentives for high performing teachers to enter or remain in schools subject to the incentives
- By altering the selection of individuals into teaching towards those who are more able to benefit from such a reward system.¹

However, as Hanover Research reviews below, it is not clear that pay-for-performance programs actually have these effects, and the use of such programs remains controversial.

VARIETIES OF PAY-FOR-PERFORMANCE MODELS

Pay-for-performance programs vary in several ways. One of these is the unit of evaluation: some pay-for-performance programs reward individual teachers for superior performance, while others reward groups — e.g., an entire school or a team of teachers within a school — if the students of the school or team meet some specified achievement goal.²

Pay-for-performance models also differ in terms of the criteria they include in their measures of teacher effectiveness, and in the weights they assign to those criteria. The three primary measures used in estimating teacher effectiveness are:

- Student achievement: Standardized test scores are the most common metric of student achievement, though other measures are required for subjects that are not included in statewide tests.

- **Classroom observations:** Trained administrators (often a principal or an external staff member) observe teachers’ practices in the classroom and assign scores according to a rubric

- **Student surveys:** These surveys gather information on students’ learning experiences in the classroom. While there are models for gathering impressions even for young children, experts generally recommend using student surveys only with older students who are more likely to understand the nature of the survey.  

Teachers’ effectiveness ratings are generated by assembling data on some combination of these measures and applying a formula that specifies the weight of each measure in determining the overall score. Based on Hanover’s review, it appears that comparatively few teacher incentive programs include student surveys as components of teacher effectiveness ratings.  

Importantly, district- or state-level teacher evaluation procedures can use different formulas for different kinds of teachers, in order to reflect appropriate evaluation criteria for teachers at a particular grade level and in a particular subject area. Evaluators may wish to weigh standardized test scores less for certain teachers if, for example, statewide standardized tests may not exist for their subjects (e.g., social studies). In that case and for some subjects, especially creative arts such as music or theater, evaluation of pencil-and-paper tests may be irrelevant or inapplicable. Thus, districts can tailor effectiveness rating systems to performance goals for different types of teachers by adjusting both the criteria included in each teacher’s formula and the weight assigned to each criterion.

**EVALUATION MODELS: VALIDITY, RELIABILITY, AND IMPACT**

However, even when a teacher rating systems employs appropriate student assessments, two questions of central importance for educators remain:

- Does the model provide an accurate and trustworthy reflection of teachers’ effectiveness in promoting student learning?
- Does basing teachers’ compensation on the model lead to increased student performance?

---


These questions reflect concern with three different features of assessment models, which may well come apart and must be addressed separately:

- **Model Validity** refers to how well the model predicts student achievement outcomes. Models with high validity are useful in identifying teachers who have particularly strong or weak effects on student achievement.

- **Model Reliability** refers to how consistently the model measures teacher effectiveness across different circumstances. Models with high reliability will correctly identify teachers as more effective or less effective across different student populations and different classroom observers.

- **Model Impact** refers to the difference that an effectiveness estimation model makes to educational practices in the environment where it is implemented — e.g., when a model is put in place in a school, district, or state does it have a positive impact on teaching practices or student test scores? Importantly, there are no guarantees that even the most valid and reliable models will be effective in improving education practices and student outcomes.

Concerning validity and reliability, research suggests that effectiveness estimates that combine multiple measures — e.g., test scores, classroom observations, and student surveys — tend to be superior in terms of both reliability and validity. However, there is no consensus about the optimal weighting of observation-based, student-survey-based, and achievement-based criteria. Indeed, a 2012 report from the Gates Foundation argues for the existence of “trade-offs, both in validity (predicting other outcomes) and reliability” among different models. Specifically, models that weigh achievement scores more heavily have greater validity in terms of predicting student achievement gains, but these models produce estimates that are less reliable, because they are more likely to show large changes in response to inter-observer variation.

In practical terms, these findings suggest that, as far as current research has established, administrators of pay-for-performance programs can use any of a number of different weighting systems for the different measures included in calculating teachers’ effectiveness scores — no one model enjoys across-the-board advantages over the others. However, administrators should understand the limitations of the particular model they choose to employ.

---


9 Ibid., p. 50-51.
**Value-Added Models**

A popular method for establishing the “student achievement” portion of the teacher-effectiveness equation is to use a value-added model (VAM). These models apply statistical procedures to indices of students’ characteristics (e.g., gender, race/ethnicity, socioeconomic status) and their past performance to project students’ likely scores on standardized tests in a given subject in a given year. Administrators can then compare these projected scores to students’ actual scores, yielding a measure of the difference that a teacher (or group of teachers) made to those students’ learning. The idea behind VAMs, then, is that “[i]n essence, the value-added method compare[s] students’ test scores in each subject area with how they would be expected to score on these tests had they been taught by the average performing team for the subject area and grade level.”\(^\text{10}\)

Value-added models are, therefore, distinct from what have been termed Student Growth Percentiles (SGPs), an alternative method for determining teacher effectiveness that tracks students’ achievement growth, but does not seek to separate teachers’ causal contributions to that growth from the contributions of other factors.\(^\text{11}\) Although SGPs have their supporters,\(^\text{12}\) experts generally agree that VAMs are a superior choice for high-stakes personnel decisions, because VAMs can, at least in principle, control for preexisting differences among students, classrooms, and schools.\(^\text{13}\)

**Validity and Reliability of VAMs**

There is a substantial amount of literature dedicated to assessing the validity and reliability of numerous VAMs. Much of this literature is highly technical, but some general lessons can be drawn that should inform the way administrators approach the use of VAMs in teacher evaluations. Below Hanover Research reviews two recent studies and then discusses their implications for decisions about VAMs at the school-district level.

A 2014 paper by Daniel Goldhaber and colleagues used a large dataset of student test scores to measure correlations among four different estimates of teacher effectiveness. That is, Goldhaber et al. sought to determine whether these models were essentially interchangeable or if, instead, using one model rather than another would produce very different classifications of teachers into effective and ineffective categories. These models (three VAMs and one Student Growth Percentile model) each included a different

\(^{10}\) Springer et al., Op. Cit., p. 372.


combination of data about students, their classrooms, and their schools. Goldhaber et al. found that three of the models were mostly interchangeable (though more so for some types of teachers than others), while the fourth produced very different rankings.

In a 2012 study, Cassandra Guarino and colleagues employed a series of simulations with different VAMs “to determine how accurately several commonly used estimation methods measure teacher performance under different scenarios that assign students to schools and teachers.” Among the conclusions the authors draw from these simulations are:

- No one method accurately captures true teacher effects in all possible assignment scenarios, although some are more robust across scenarios than others
- The probability that an above-average teacher can be misclassified as below average can be fairly high

Studies like these two hold two obvious lessons for administrators faced with decisions about whether and how to employ VAMs in teacher evaluations:

- Since value-added estimates can be highly sensitive to small changes in educational context, districts should exercise caution both in drawing conclusions about teacher effectiveness from these estimates, and in basing high-stakes personnel decisions on those conclusions.
- Determining which VAM is most appropriate for a given evaluation context can require significant technical expertise. Before attempting to implement a VAM-based assessment model, districts should ensure that staff members (or outside consultants) with the requisite expertise are available to advise administrators on technical aspects of VAM design and implementation.

Congruent with the first of these lessons, Guarino et al. are skeptical about the use of the VAMs they consider for high-stakes personnel decisions: “[t]he probability that a teacher in the top or bottom quintiles of the quality distribution is correctly identified as such is lower than would be desirable for purposes that link performance to rewards or sanctions.” The second lesson is echoed in a recent position paper on VAMs from the American Statistical Association: “VAMs are complex statistical models, and high-level statistical expertise is needed to develop the models and interpret their results.”

---

15 Bulleted text taken verbatim from: Ibid., p. 4.
16 Ibid., p. 4.
SECTION II: IMPACTS OF PAY-FOR-PERFORMANCE AND VALUE-ADDED MODELS

In this section, Hanover Research reviews current research on the effectiveness of pay-for-performance and value-added model programs in motivating teachers to improve their teaching practices and raise student performance. Hanover Research begins with a summary of research on a variety of pay-for-performance programs, followed by a review of the latest findings on the effectiveness of pay-for-performance programs that rely on VAMs for estimates of teacher effectiveness.

EFFECTIVENESS OF PAY-FOR-PERFORMANCE PROGRAMS

There is some evidence to suggest that teacher performance-based-pay programs can increase student performance. For instance, a 2007 study by David Figlio and Lawrence Kenny of the National Bureau of Economic Research found that districts with individual financial incentives for teachers experienced greater improvements in student test scores than did districts without such incentives. This effect was only observed in low- to mid-socioeconomic status classrooms; the incentives had no effect in highly advantaged classrooms.\(^{18}\)

Importantly, however, Figlio and Kenny’s study collected very little information about the incentive programs used. So, while these results suggest that financial incentives can have positive effects on student outcomes, they inform little about what specific program features might drive this effect. Additionally, pertinent to the topic of this report, Figlio and Kenny’s study does not address the role of VAMs in these incentive programs.

An earlier study conducted by Firestone and Pennell presents a more skeptical view of teacher incentive programs.\(^ {19} \) Rather than focusing on student outcomes, Firestone and Pennell examined teachers’ attitudes towards pay-for-performance programs and the effects that competitive incentive programs have on teachers’ motivation to work with one another. In the face of these individual-based incentive programs, educators tend to find fault with the measures of performance. For instance, when rewards are based on observation, teachers may claim observer bias; when rewards are based on standardized test results, teachers may argue that scores are influenced by factors outside of their control, such as students’ socioeconomic status. These concerns regarding the fairness of evaluation can cause teacher performance to suffer.\(^ {20} \) Similarly, selective incentive

---


\(^{20}\) Ibid., p. 513.
programs increase competition among teachers for rewards and may, therefore, negatively influence collaboration.\textsuperscript{21}

Results from the District of Columbia Public Schools’ (DCPS) IMPACT program exemplify the mixed nature of evidence on the relationship between teacher incentives and educational outcomes. The IMPACT program, which began in 2009, combines student achievement data with observational scores and other measures to assign teachers to one of four effectiveness categories. Unlike most other teacher incentive programs, the IMPACT program includes not only incentives for highly effective teaching, but also disincentives for ineffective teaching: teachers with low effectiveness scores may face dismissal.\textsuperscript{22}

In 2013, Thomas Dee of Stanford University and James Wyckoff of the University of Virginia analyzed the effects of the IMPACT program on students’ and teachers’ outcomes.\textsuperscript{23} These results are of interest in the context of the present report because, although test scores do not solely determine estimates of teachers’ performance under the IMPACT system, a value-added model for scores on DC’s Comprehensive Assessment System (CAS) tests is nevertheless a key component of the evaluation system.

Dee and Wyckoff found that the IMPACT program had positive effects on teacher’s performance, but only for a small subset of teachers. IMPACT’s incentives had little or no effect on teachers in the middle of the performance distribution, but two groups of teachers did show performance improvements. These were teachers who earned performance bonuses in the previous year, and teachers whose scores were close to the threshold for sanctions (i.e., dismissal). Dee and Wyckoff suggest that this is because these teachers had the greatest incentives to improve: bonus-earning teachers were eligible for even larger benefits if they received another high rating, and borderline-effective teachers risked dismissal if they did not raise their scores.

However, it is not clear whether these effects translated into gains in students’ achievement. Since teachers’ performance ratings are a composite of student test scores and other measures, the “performance improvements” seen in teachers’ scores may not reflect changes in student outcomes, but rather changes in other measures, such as observation scores. Moreover, because the IMPACT initiative did not include a control group, it would be difficult to determine in any case whether any student achievement gains could be attributed to the effects of the incentive program.

**Research on Value-Added Models: Effectiveness and Best Practices**

In the last few years, several large-scale experimental or quasi-experimental studies have attempted to isolate the effects of implementing an incentive model on teacher practices and student achievement. Most of these studies have failed to identify any effects on these outcomes. In particular, three randomized controlled trials in the last several years

\textsuperscript{21} Ibid., p. 514.
\textsuperscript{22} Dee and Wyckoff, Op. Cit.
\textsuperscript{23} Ibid.
have yielded null results. Below Hanover Research provides brief descriptions of these studies (more complete information on the features of the incentive program investigated in each study may be found in Figure 3.1 in Section III of this report), then discusses particular findings that provide useful information for the design and implementation of teacher pay-for-performance programs. The three studies are:

- **Project on Incentives in Teaching (POINT), TN:** From 2006 to 2009 middle schools in the Metropolitan Nashville School System distributed rewards of $5,000 to $15,000 annually to individual math teachers based on value-added measures of effectiveness. A team of researchers from Vanderbilt University and the RAND Corporation evaluated the POINT program in 2010, and found that the incentives had virtually no effect on either students’ performance or teachers’ practices.

- **Round Rock Independent School District (RRISD), TX:** The Vanderbilt/RAND group also studied the effects of RRISD’s group-based pay-for-performance program, which distributes financial rewards teams of teachers based on VAM measures (profiled in Section III). After comparing test scores among students whose teachers did and did not participate in the incentive program, the researchers concluded that “a]nalysis of student achievement outcomes reveals no overall intervention effect in any subject area across the 2 years of the experiment.” As in the POINT evaluation, the researchers also found no effects on teachers.

- **Schoolwide Performance Bonus Program (SPBP), New York City Public Schools, NY:** In 2007, the New York City Department of Education (NYCDOE) implemented a group-based teacher incentive program in approximately 200 underperforming New York City schools. Two separate analyses of the results of the program found no effects on teacher practices or attitudes, or on student outcomes. Following the release of these studies, New York’s Department of Education terminated the bonus program in 2011.

Although these studies suggest that teacher incentive programs are not the perfect cure for improving student achievement, a close examination of these incentive programs can still identify factors that may have contributed to the outcomes — or lack thereof — in the

---

24 In fact, there is a fourth such study, conducted in suburban Chicago in 2010-2011, that investigates the impact of pay-for-performance teacher incentives from the perspective of behavioral economic theories of loss aversion. The results are intriguing and suggest a possible effect of teacher incentives on student achievement. However, the study employs a rather artificial design that would be difficult to implement in a non-experimental setting. We omit discussion of this study in this section, but a description may be found in Appendix A.


study. Below Hanover Research discusses two such factors and their relevance to designing and implementing pay-for-performance programs below.

**Teachers’ Program Understanding**

A common finding in studies of the ineffectual pay-for-performance programs discussed above is that teachers (and sometimes principals) often find the procedures for generating value-added scores confusing, and may not fully understand the criteria by which bonuses are awarded. For example, 54 percent of teachers who participated in RRISD’s incentive program reported that they did not have a clear understanding of the criteria for earning a bonus, and a substantial number (39 percent) remained confused even after reading an FAQ document describing the procedure for awarding bonuses. 29

In New York City’s program, teacher understanding was also low, especially in the first year of the program. While program administrators managed to improve teachers’ understanding by requiring additional training about evaluation and reward procedures, communication issues and misconceptions about the program persisted. 30 Figure 2.1 shows teachers’ understanding of several program features in year three of the initiative: while a majority of teachers reported understanding most program features at least somewhat well, a considerable number still lacked an adequate grasp of the program.

**Figure 2.1: Percentage of Teachers Reporting Level of Understanding of SPBP (NY) Elements**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Not Very Well</th>
<th>Somewhat Well</th>
<th>Very Well</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which staff are eligible to receive a share of the schoolwide bonus</td>
<td>27%</td>
<td>45%</td>
<td>28%</td>
</tr>
<tr>
<td>Criteria for a school to receive a full bonus</td>
<td>29%</td>
<td>48%</td>
<td>23%</td>
</tr>
<tr>
<td>How a school compensation committee decides on a distribution plan for the bonus</td>
<td>34%</td>
<td>40%</td>
<td>26%</td>
</tr>
<tr>
<td>The target our school needs to reach in order to earn a bonus this year</td>
<td>34%</td>
<td>43%</td>
<td>23%</td>
</tr>
<tr>
<td>The amount of money our school would receive if we met 100% of our target</td>
<td>41%</td>
<td>44%</td>
<td>15%</td>
</tr>
<tr>
<td>The source of funding for this program</td>
<td>56%</td>
<td>34%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Source: Marsh et al. 31

Poor program understanding among teachers presents at least two threats to the effectiveness of a pay-for-performance initiative. First, if teachers do not understand how their performance is measured or how performance is linked to rewards, they may be unable to modify their practices to improve performance and obtain a reward, even if they
are motivated to do so.\textsuperscript{32} Second, misconceptions about the method for awarding bonuses can reduce teacher buy-in if these misconceptions create the impression that the method is unfair.\textsuperscript{33}

These considerations, together with the prevalence of communication issues revealed in the research literature, underscore the importance of developing effective communication practices for teacher incentive programs. In particular, the pattern of communication difficulties documented in the research literature suggests the following best practices for program administrators:

- Require all teachers and principals to attend orientation sessions at the outset of the program
- Develop clear and non-technical written materials, with clear instructions about where to direct further questions
- Ensure that all personnel who are likely to field questions from teachers (e.g., principals, evaluation specialists, etc.) are especially well-versed in program features
- Gather detailed feedback from teachers about their attitudes towards and knowledge of program features. Notably, Marsh et al. found that some teachers reported a high level of understanding in response to surveys, but demonstrated misconceptions about the program in other contexts.\textsuperscript{34} Informal assessment tools that allow teachers and administrators to check teachers’ understanding of program structure may therefore be helpful.

**Prior Context and School Characteristics**

Pay-for-performance models may not provide the same incentives in all teaching environments. For instance, New York’s incentive program was restricted to schools that were already underperforming, and researchers have noted this as a possible reason for the failure of incentives to make a difference.\textsuperscript{35} Teachers at these schools likely already faced significant pressure to increase performance, and if these teachers were already highly motivated to improve student achievement, additional financial incentives would have little impact on their behavior. While future research is needed to test this hypothesis, the lesson for administrators of teacher incentive programs is clear: characteristics of individual schools and teacher teams can make a difference to the efficacy of teacher incentive programs. Therefore, administrators should consider how the demographics, history, and culture of schools in the district might affect teachers’ responsiveness to incentives.

\textsuperscript{32} Ibid., p. 115.
\textsuperscript{34} Marsh et al., Op. Cit., p. 116.
**SECTION III: DISTRICT PROFILES**

School districts around the country have implemented a wide variety of teacher pay-for-performance systems using VAMs; in this section, Hanover Research presents detailed profiles of Round Rock Independent Schools and Harrison School District and a summary of several other programs in the table below.

**Figure 3.1: Key Features of a Sample of pay-for-performance Programs**

<table>
<thead>
<tr>
<th>DISTRICT</th>
<th>INDIVIDUAL OR GROUP INCENTIVES?</th>
<th>EFFECTIVENESS ESTIMATION MODEL</th>
<th>REWARD STRUCTURE</th>
<th>QUALIFYING CRITERIA FOR BONUS</th>
<th>BONUS VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round Rock Independent School District, TX</td>
<td>Group (teacher teams)</td>
<td>Achievement (VAM) &lt;br&gt; o Subjects: Math, ELA, Science, Social Studies</td>
<td>All or nothing</td>
<td>Top 1/3 of teacher teams</td>
<td>$6,000 per teacher (tested subjects) &lt;br&gt; $3,800 - $5,500 (non-tested subjects)</td>
</tr>
<tr>
<td>Washington, DC</td>
<td>Individual</td>
<td>50% Achievement &lt;br&gt; o Subjects: Math, ELA, Science, Health &lt;br&gt; 50% Observation, etc.</td>
<td>Increased rewards for multi-year excellence</td>
<td>“Highly Effective (HE)” rating</td>
<td>Up to $5,000 bonus for “HE” rating &lt;br&gt; Base pay increases of up to $7,000 for two consecutive years of “HE” ratings</td>
</tr>
<tr>
<td>Nashville, TN</td>
<td>Individual</td>
<td>Achievement (VAM) &lt;br&gt; o Subject: Math</td>
<td>Increasing rewards at 3 performance tiers</td>
<td>80th percentile for 1st tier &lt;br&gt; 85th percentile for 2nd tier &lt;br&gt; 90th percentile for 3rd tier</td>
<td>$5,000 (1st tier) &lt;br&gt; $10,000 (2nd tier) &lt;br&gt; $15,000 (3rd tier)</td>
</tr>
<tr>
<td>Harrison School District Two, CO</td>
<td>Individual</td>
<td>50% Achievement (VAM) &lt;br&gt; o Subjects: Varies by Teacher &lt;br&gt; 50% Observation</td>
<td>Base salary increases as teacher is promoted through 9 effectiveness tiers</td>
<td>Achieving target effectiveness score triggers promotion to next tier</td>
<td>Base salary raise of $6,000 - $10,000 for each promotion</td>
</tr>
<tr>
<td>New York City Public Schools, NY*</td>
<td>Group (schoolwide)</td>
<td>Improvement on composite of test scores, school environment, and academic progress</td>
<td>Increasing rewards at 2 performance tiers</td>
<td>School must meet 75% (1st tier) or 100% (2nd tier) of specified targets</td>
<td>Qualifying schools awarded $1,500 - $3,000 per teacher &lt;br&gt; o Each school determined internal distribution</td>
</tr>
<tr>
<td>Chicago Heights, IL†</td>
<td>Individual</td>
<td>Achievement (VAM) &lt;br&gt; o Subject: Math</td>
<td>Increases linearly with effectiveness scores up to maximum bonus</td>
<td>All but lowest 10%</td>
<td>Up to $8,000</td>
</tr>
</tbody>
</table>

Source: Compiled by Hanover Research from sources cited in this report
*Not active, ended by New York Department of Education  †Not active, experimental manipulation only (see Appendix B)
Harrison School District Two, Colorado

Harrison School District, located in Colorado Springs, Colorado, comprises 25 schools with 10,000 enrollees, 36 70 percent of whom qualify for free or reduced lunch.37 In 2010, Harrison launched “what is arguably the boldest pay-for-performance plan in the country.”38 In a 2012 report sponsored by the Fordham Institute, an education policy think-tank, former Harrison superintendent Mike Miles (now superintendent of the Dallas Independent School District, with plans to create a similar compensation plan in that district39) provides a detailed account of the structure and implementation of Harrison’s teacher compensation plan.40 Hanover’s profile of Harrison’s initiative reviews this and other reports on the Harrison model, and draws out best practices for introducing a performance-based teacher compensation plan.

Program Outline

Unlike many of the pay-for-performance models discussed above, Harrison’s approach does not offer performance-based compensation in the form of lump-sum bonuses. Rather, Harrison uses a performance- and achievement-based assessment model to determine all promotions and salary increases. In other words, Harrison has eliminated all teacher salary adjustments that are not performance-based, including annual raises, cost-of-living raises, and raises based on academic or professional credentials.41

Harrison’s evaluation and compensation plan assigns teachers to one of nine “effectiveness levels,” each corresponding to a salary level (Figure 3.2). Teachers earn raises by achieving target scores on annual performance evaluations, which moves them into a higher tier/salary level.

Under Harrison’s performance evaluation formula, achievement-based and observation-based criteria each count for 50 percent of the overall score, with each category broken down into several sub-criteria. The achievement portion of the score reflects students’ performance on a variety of assessments ranging from state standardized tests to district-level exams to art projects. These assessments are used in different combinations to measure effectiveness for teachers at different grade levels and in different subject areas.42 Thus, Harrison’s evaluation metric exemplifies one of the best practices reviewed in Section I, because it combines multiple measures to create assessments of teachers’ performance that are tailored to the different responsibilities of different teachers.

41 Ibid., p. 6.
42 Ibid., p. 7-11.
Figure 3.2: Harrison School District Effectiveness Levels and Salary Levels

<table>
<thead>
<tr>
<th></th>
<th>Novice</th>
<th>Progressing</th>
<th>Proficient</th>
<th>Exemplary</th>
<th>Master</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Salary</td>
<td>35k</td>
<td>38k</td>
<td>40k/44k</td>
<td>48k</td>
<td>54k</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>60k</td>
<td>70k</td>
<td>80k</td>
</tr>
</tbody>
</table>

Source: Miles, 2012.43

**Program Budget**

The Harrison School District was able to enact its pay-for-performance model despite recession-driven budget cuts. Miles notes that Harrison minimized startup costs by developing its incentive plan almost entirely in-house, though the district also secured a two-year, $800,000 grant from the Daniels Fund to support program development and implementation.44 Miles characterizes the budgetary impact of the program as follows:

- **Startup Costs: $700,000**
  - *Initial salary adjustments* Harrison assigned teachers to effectiveness levels as soon as the new compensation program launched, and adjusted salaries accordingly. Eighty percent of teachers received salary increases, for a total cost of $300,000 over the previous year’s salaries.
  - *Assessment development and scoring.* In order to develop high-quality student assessments that could serve as the basis for measures of teaching effectiveness, Harrison added three staff members to its five-person Curriculum and Assessment Department, and enlisted consultants to design district-wide exams. Together, the additional staff and consulting services cost $400,000.

- **Program Maintenance Costs: Cost-neutral in the medium term**
  - Although Harrison will distribute substantial pay raises to some teachers each year, these costs are offset by the elimination of automatic pay raises. The Harrison compensation plan allows for teachers to move quickly from the Novice level to Proficient I, with promotions (and raises) becoming progressively more difficult thereafter. Miles notes that “[a]s long as the plan stays rigorous, with an estimated 20 to 25 percent of teachers promoted each year, the plan is financially sustainable.”45
  - It is nevertheless a feature of the Harrison model that total salary commitments could increase significantly in the long term, if enough teachers qualify for the higher effectiveness tiers. But since salary increases are tied to student achievement, this can happen only when students are achieving significantly higher scores. Miles calls this “a tradeoff we welcome.”46

---

43 Ibid., p. 7.
44 Ibid., p. 18.
46 Ibid., p. 19.
RECEPTION AND IMPACT

Although Harrison’s pay-for-performance model has received widespread recognition,\textsuperscript{47} reception of the compensation plan among teachers has been mixed. However, annual teacher surveys administered by the district show that the program gained popularity over the first two years of its implementation, with teachers less likely to say they disliked the program and more likely to approve of pay-for-performance models generally in the second year.\textsuperscript{48} However, some teachers have expressed doubts that Harrison’s evaluation model accurately measures their teaching performance. In response to such worries, Miles stressed the importance of maintaining flexibility in the way different components are weighted in evaluations of teachers at a given grade level and subject area.\textsuperscript{49}

It is not clear whether Harrison’s pay-for-performance plan is having any effect on student outcomes. The scores of Harrison students on state standardized tests have risen over the last several years — and at a faster rate than most other districts in the state — but this appears to be a continuation of a trend that began several years before implementation of the new compensation plan.\textsuperscript{50}

LESSONS LEARNED

Mike Miles summarizes the following “lessons learned” from Harrison’s experience creating and implementing a performance-based compensation model. These principles outline concrete steps administrators can take at multiple junctures to ensure a smooth transition to a pay-for-performance system. In addition to these lessons learned, Hanover Research provides a series of “key details” of the Harrison plan (found in Appendix B), which give a fuller picture of how Harrison managed various logistical issues introduced by the transition to a performance-based system.

- Draft a concept paper and framework; then get input.
  - While this may seem to be a “top down” approach, teachers want to know what they will be held accountable for.

\textsuperscript{50} Ibid.

- Over-communicate with principals and other building leaders.
  - Most teachers turn to their principals first, and some Harrison principals did not have enough information about certain aspects of the plan. (Cf. subsection “Teachers’ Program Understanding” in Section II above.)

- Pilot any new assessment tools for a year.
  - This gives teachers and the district a chance to work out the kinks.

- Ensure the curricula are well-written and finalized before trying to create aligned tests.

- Share any new assessment tools with teachers at least two months prior to the date they will be administered.

- Build in assessment and scoring days.

- Create a focus group with representatives from each school.
  - Hold regular meetings to promote teacher buy-in and information flow between teachers and administrators.

- Standardize the processes for creating, delivering, administering, and scoring assessments.

- Try something, but promise to address problems and make improvements along the way.
  - If a district waited until it created a perfect pay-for-performance plan, nothing would ever change.51

### ROUND ROCK INDEPENDENT SCHOOL DISTRICT, TEXAS

Round Rock Independent School District (RRISD) serves the area around Austin, Texas, and enrolls approximately 43,000 students across 50 schools.52 In 2010, RRISD, in collaboration with the National Center on Performance Incentives53 and with funding from the U.S. Department of Education’s Teacher Incentive Fund (TIF),54 implemented a group-based pay-for-performance initiative. The implementation of RRISD’s program exhibits a number of best practices to building an effective pay-for-performance initiative. After a brief outline of the program, we describe how RRISD addressed a common problem for pay-for-performance systems and promoted communication and teacher understanding.

#### PROGRAM OUTLINE

RRISD’s incentive program rewards teams of teachers who have positive effects on students’ scores on tests in reading/language arts, mathematics, science, and social studies.

---

53 Ibid., p. 370.
RRISD’s program used a value-added model that estimated each team’s impact on student performance. The top one-third of teacher teams, as determined by value added to students’ test scores, received bonuses of approximately $5,500 per teacher.

**Addressing Communication and Teacher Understanding Issues**

An external review conducted two years into the program found that the communication difficulties surrounding program mechanics discussed in the previous section manifested in the RRISD program as well. The reviewers note that the communication difficulties were quite pronounced in the first semester of the program. In response, administrators made three changes to the program:

- Principals were brought into the program operations at a much more involved level, which increased their knowledge of the program and their abilities to communicate specifics to their campus teams.

- The district-level program staff held a series of trainings on the program components and the changes for Year 2, traveling to each campus to outline the specifics of each program requirement. Changes in the delivery of the information to include more hands-on activities and small-group discussions increased the ability of participants to understand and retain the information.

- Communication processes have improved, and a monthly newsletter is now circulated that details program information, changes, training opportunities, meeting updates, etc.

By the second semester of the program — after these changes were implemented — focus groups with teachers established that teachers had a high level of program understanding.

---

55 Ibid., pp. 4-5.
56 Bulleted text taken nearly verbatim from: Ibid., pp 4-5.
57 Ibid., p. 4.
APPENDIX A: TEACHER INCENTIVES AND LOSS AVERSION

A 2012 study from a team of researchers, led by economist Roland Fryer of Harvard University, found that the framing of an incentive program was an important mediator of its impact on student performance. The researchers randomly assigned teachers in a suburban Chicago school district into “Gain” and “Loss” groups. The “Gain” group was enrolled in a traditional incentive program that offered bonuses of up to $8,000 based on student achievement gains, while the “Loss” group received a lump sum of $4,000 at the beginning of the year, with the understanding that they could earn further rewards to up to $4,000 if their students performed above average that year, but they were required to return some portion of the original $4,000 if their students’ performance was below average. Importantly, then, “Gain’ and ‘Loss’ teachers received identical net payments for a given level of performance. The only difference is the timing and framing of the rewards.”

Fryer et al. found no improvements in scores among the students of “Gain” participants, but significant improvements among those of “Loss” participants: math scores among “Gain” teachers’ students rose by .22 standard deviations, or 6.8 percentage points. The researchers found similar patterns for reading scores, though they express less confidence in the methodology employed in that part of the study. These results are consistent with a large body of research demonstrating that incentives framed as losses are often more motivating than identical incentives framed as gains. As Fryer et al. put it: “giving back the cash advance feels very different to people than not getting paid a bonus, even though ultimately the same financial end result is achieved.”

While Fryer et al.’s study is preliminary and its design untested in more naturalistic environments, the results suggest that the way an incentive program is implemented, and, in particular, the timing and framing of rewards, can make the difference between an incentive program that boosts student achievement and one that does not.

---

59 Ibid., p. 10
60 Ibid., p. 13
APPENDIX B: KEY DETAILS OF HARRISON COUNTY SCHOOL DISTRICT’S PAY-FOR-PERFORMANCE PLAN

TEACHER PLACEMENT & MOBILITY

INITIAL PLACEMENT

New, first-year teachers start at the Novice effectiveness level. New teachers who have teaching experience may be placed at Novice, Progressing I, Progressing II-a, Progressing II-b, or Proficient I. Placement is made by the Human Resources Department in collaboration with the principal and with input from the superintendent and the Department of School Supervision and Leadership (SSL). The superintendent, with input from HR, may place a new teacher at a higher level if there is evidence that the teacher has produced exceptional student-achievement results and is an exemplary teacher. These exceptions are made on a case-by-case basis.

TENURE

Until the passage of teacher effectiveness legislation in May 2010 (S.B. 191), the Harrison Plan did not change or affect tenure. Teachers would still have achieved “non-probationary” status once they began their fourth year in the district. Now, based on revisions to Colorado law, tenure (or non-probationary status) will be granted after three consecutive years of “effective” teaching. The district aligned its definition of effectiveness with the law, defining an “effective” teacher as one who receives an overall evaluation of Proficient I or higher.

Probationary teachers who do not meet the district’s standards may be non-renewed. Non-probationary teachers who are not performing at the Proficient level, according to the district’s evaluation system, may undergo remediation and, if necessary, be terminated.

ADVANCEMENT TO THE NEXT EFFECTIVENESS LEVEL

After initial placement, teachers advance from one effectiveness level to the next without skipping a level. However, there is no minimum number of years that a person must remain at a certain level.

In exceptional situations, the superintendent, after consultation with a review board, may adjust upward a teacher’s placement on the plan’s scale. (The superintendent determines the makeup of the review board and convenes the board as needed.) Adjustment of a teacher’s placement on the scale in this manner may be done only once in a teacher’s career in the Harrison District.
MOVEMENT TO A LOWER LEVEL

A teacher may be moved to a lower level after receiving a lower evaluation rating for three consecutive years. Starting in the 2014-2015 school year, a teacher may be moved to a lower level after two consecutive years of lower performance. The teacher will remain at that lower level for at least one year and will receive the salary associated with that level (except that the salary of a non-probationary teacher currently employed full-time by the district may not be lower than his 2009-2010 salary).

TRANSFER TO ANOTHER SCHOOL IN THE DISTRICT

Proficient II and higher teachers may be transferred to a school that requires more skilled teachers.

EVALUATIONS

TIMELINE FOR EVALUATIONS

Under the Harrison Plan (and S.B. 191), all teachers receive a summative evaluation every year. School administrators usually conduct summative evaluations toward the end of the school year, during April and May.

Beginning in the 2011-2012 school year, school administrators may also conduct summative evaluations at the end of the first semester. These mid-year evaluations are reserved for teachers who are at Progressing II or higher and are likely to be advanced to the next effectiveness level.

For both the end-of-year and mid-year evaluations, the review must take into account the teacher’s performance over at least three academic quarters. It must also take into account achievement data compiled over four academic quarters.

APPEAL/REVIEW OF EVALUATION RATING

There is no appeal of one’s evaluation rating. However, if the performance score and the achievement score differ by more than one level, the teacher may request a review of her evaluation.

Upon such a request, the district’s Review Team reviews the teacher’s performance, student achievement data, and the principal’s process for establishing the teacher’s rating. The Review Team comprises members of the SSL Department and Curriculum Department as determined by the superintendent. The Review Team makes a rating recommendation to the superintendent. The superintendent then makes the final determination after considering the recommendation of the Review Team and receiving input from HR.
**Remediation/ Improvement Plans**

Teachers on an improvement or remediation plan must successfully complete the plan before they may be considered for advancement. They must be evaluated for at least one year after the date of the end of the improvement or remediation plan before being given a summative evaluation that could result in a promotion to the next level.

**Student Teachers or Interns**

Teachers are held accountable for their students’ scores even if they accept a student teacher or intern.

**Quotas**

There is no “quota” for the percentage of teachers who can be at each effectiveness level.

**Effect of Student Attendance on Scores**

The scores of “chronically absent students” — students who are absent more than 25 percent of the time — are not included in a teacher’s data set and do not count against a teacher.

Students must also meet the eligibility requirement for their scores to be counted. For example, the scores of a student who has not been assigned to a particular teacher for at least six weeks prior to a semester or quarterly exam will not count for or against a teacher’s data.
PROJECT EVALUATION FORM

Hanover Research is committed to providing a work product that meets or exceeds partner expectations. In keeping with that goal, we would like to hear your opinions regarding our reports. Feedback is critically important and serves as the strongest mechanism by which we tailor our research to your organization. When you have had a chance to evaluate this report, please take a moment to fill out the following questionnaire.


CAVEAT

The publisher and authors have used their best efforts in preparing this brief. The publisher and authors make no representations or warranties with respect to the accuracy or completeness of the contents of this brief and specifically disclaim any implied warranties of fitness for a particular purpose. There are no warranties which extend beyond the descriptions contained in this paragraph. No warranty may be created or extended by representatives of Hanover Research or its marketing materials. The accuracy and completeness of the information provided herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every partner. Neither the publisher nor the authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Moreover, Hanover Research is not engaged in rendering legal, accounting, or other professional services. Partners requiring such services are advised to consult an appropriate professional.