

Denver

Highlighted below are CTAC's analyses, findings and recommendations. All of the recommendations are crafted to meet the standard of increasing the manageability, fairness and sustainability of Pay for Performance. The report also examines the national implications of Denver's Pay for Performance pilot. The issues are complex and multi-faceted, and are discussed in full detail in the chapters of the report.

A. Primary Findings

Impact on Student Achievement

- At all three academic levels—elementary, middle, and high school—higher mean student achievement in the pilot schools is positively associated with the highest quality objectives. Students whose teachers had excellent objectives, based on a four-level rubric developed by CTAC, achieved higher mean scores than students whose teachers' objectives were scored lower on the rubric. This holds true on most tests of the *Iowa Test of Basic Skills (ITBS)* and the *Colorado Student Assessment Program (CSAP)*.

At the elementary school level, students of teachers with excellent objectives (rubric level 4) had significantly higher mean normal curve equivalent (NCE) scores on the *ITBS* Reading, *ITBS* Language, and *CSAP* Math tests than students of teachers with lower rubric levels.

At the middle school level, students of teachers with level 4 objectives had significantly higher mean scores than students of teachers with level 3 and level 2 objectives on the *CSAP* Math test.

At the high school level, there were significantly higher mean scores on the *ITBS* Reading and *CSAP* Writing tests at Manual High School and on the *CSAP* Writing test at Thomas Jefferson High School for students whose teachers had level 4 objectives.

Six other tests (one elementary, three middle school, and two high school level) show a positive relationship between highest quality objectives and higher achievement level that is not statistically significant.

- Pilot students were compared to control students, estimating the change in mean NCE scores over time (e.g., from the baseline year through the end of the pilot) on three tests of the *ITBS* and three tests of the *CSAP*. Two-stage hierarchical linear modeling (HLM) was used to account for correlation between observations and to adjust for differences in school and student characteristics. The use of NCEs makes it possible to determine whether students are performing better than expected based on previous scores—attaining more than an expected year of growth—or not.

The effect of the pilot varies by level of school:

The pilot elementary school mean NCE scores declined on the *ITBS* Math, *CSAP* Reading and *CSAP* Math tests. The control school mean NCE scores declined on the *ITBS* Math test and increased on the *CSAP* Writing test. These results were statistically significant. The performance of the pilot students was lower than the controls on the tests except on the *ITBS* Language.

The pilot middle school students performed significantly higher than the controls on the *ITBS* Reading, *CSAP* Writing, and *CSAP* Math tests. The average NCE scores of middle school pilot students increased significantly over time (i.e., attained more than a year's expected growth) on the *ITBS* Reading, *CSAP* Writing, and *CSAP* Math tests. The controls experienced statistically significant declines in mean NCE scores on the three *ITBS* tests over the course of the pilot and statistically significant increases on the three *CSAP* tests. On the *ITBS* Language test, both the pilots and controls decreased by comparable amounts.

The high school pilots and controls experienced statistically significant increases in mean NCE scores over time on most tests. Manual High School students had significantly higher increases than the control school students on the *ITBS* Language and Math tests. Thomas Jefferson High School students performed significantly higher than the control school students on the *ITBS* Language, *ITBS* Math and the *CSAP* Reading tests and significantly lower on the *ITBS* Reading test.

- Meeting two objectives is positively associated with higher mean achievement scores.

At the elementary schools, the students of teachers who met two objectives had significantly higher mean NCE scores on all six tests than students of teachers who met one objective.

At the middle schools, meeting one or two objectives was associated with significantly higher mean NCE scores than meeting no objectives on the *ITBS* Reading and Language tests. Similar positive relationships were seen on the *ITBS* Math and *CSAP* Math tests but they are not statistically significant.

At the high schools, the students of teachers who met two objectives (at both Manual and Thomas Jefferson) had significantly higher mean NCE scores than students of teachers who met one objective or no objectives on the *ITBS* Reading test. Similar positive relationships are found on the *ITBS* Math, *CSAP* Writing, and *CSAP* Math tests at Manual High School and on *ITBS* Math, *CSAP* Reading, and *CSAP* Math at Thomas Jefferson High School but they are not statistically significant.

- Student achievement rises as length of teacher participation in the pilot rises.

Elementary students whose teacher had been in the pilot for two, three, and four years had mean *ITBS* Reading scores 0.8, 1.3, and 2.2 NCEs higher than students of one-year teachers. Elementary students of four-year teachers also had mean *ITBS* Math scores significantly higher than those of one-year teachers.

Middle school students of two-year teachers scored 2 NCEs higher on average and students of three-year teachers scored 3.2 NCEs higher than students of one-year teachers.

High school students of two-year pilot teachers scored higher on *ITBS* Reading, *ITBS* Math, and *CSAP* Reading at both pilot high schools, but the difference was only statistically significant for the Thomas Jefferson *ITBS* Reading exam.

- The pilot has been the catalyst for developing a fundamentally new compensation plan for teachers in Denver which is based, in part, on student achievement.

Impact of Objectives

- The percent of teachers who developed objectives that were rated at the two highest levels of the rubric increased steadily over the course of the pilot. The particular improvement in the objectives in the final year of the pilot is largely attributable to greater attention to learning content in the objectives. By the close of the pilot, 28% of the objectives were at level four (excellent) and 44% were at level three (adequate).
- There is a significant increase in the quality of the objectives as the number of years a classroom teacher participated in the pilot increases.
- Teachers met their objectives at a high rate. The data collected by the district over the four years of the pilot show that from 89% to 93% of the teachers met one or more objectives and were awarded additional compensation.

- As teachers gained more years of experience in the pilot, their ability to meet their objectives also increased significantly. One-year pilot participants met 89% of their objectives. The success rate rose to 98% for four-year participants.

- Educational background and years of experience in the Denver Public Schools are related to whether classroom teachers met their objectives.

Certified teachers met 92% of their objectives, while teachers participating in the Teacher-in-Residence alternative certification program met 83% of their objectives.

First year teachers met 86% of their objectives, while teachers with two or more years of experience met 92% of their objectives.

Teachers with 15 or more years of experience in the Denver Public Schools met their objectives at lower rates (85%) than teachers with fewer than four years (95%), four to 10 years (90%), or 11 to 14 years (95%) of experience.

- There are similarities between pilot school teacher objectives and control school teacher goals. However, the control school teachers make less use of baseline data. Some of the similarities are attributable to the pilot's objective setting protocol being in full or partial use in nearly one-third (32%) of the control school goals reviewed in 2002-2003.
- The objectives and their learning content are not included as part of the strategies in the school improvement plans.

Perceptions of Participants and Other Parties

- The pilot has significantly increased the school and district focus on student achievement. This focus has increased with each succeeding year of pilot implementation.
- Teachers indicate that they have greater access to student achievement data and that they use the data more effectively, particularly baseline data, to establish growth expectations, to focus earlier on students who may need more assistance and to monitor progress.
- Most pilot teachers do not attribute changes in their core classroom instructional practices to the pilot. Teachers indicate that they did not receive a mandate to make such changes.
- Most teachers feel that cooperation among teachers has improved or stayed the same at the pilot schools.
- Pilot teachers are less fearful of pay for performance than control school teachers. By the end of the pilot, pilot participants were more likely to offer suggestions for improvement than to indicate that pay for performance was not viable.
- Pilot teachers continued throughout the pilot to raise issues of fairness and trust in the objective setting and review process. However, they believe that it is possible to set fair objectives.
- The quality of interaction between the principals and teachers is pivotal to the implementation of Pay for Performance at the school sites. Teachers in the pilot schools believe that there are inconsistencies from school to school in how principals review and analyze progress on teacher objectives. Principals indicate that there is a lack of clarity regarding their role and authority and a need for targeted professional development.
- Parents indicate that a teacher's contribution to student achievement should be rewarded in financial terms.

- The vast majority of parents (94%) and teachers (93%) feel that more than one measure of student achievement should be used to determine teacher performance.
- Participants value the training they received, but express a need for more professional development based on the specific student achievement levels in the individual schools and classrooms and the instructional challenges of meeting objectives.

Institutional Factors

- The DPS/DCTA collaboration on behalf of student achievement has been significant. This collaboration has been pivotal to the development of Pay for Performance despite changes in district leadership and structure.
- Pay for Performance has enabled issues which have adversely affected district progress, sometimes for many years, to be put on center stage. Operating in a climate protected by external supporters and internal reformers, the pilot provided a vehicle for problems to be discussed, analyzed and acted upon. These actions have helped the district to develop an increased capacity to make mid-course corrections.
- Teachers and principals were provided with multiple opportunities through the study to influence the course of the pilot. For many, this was a marked and positive departure from past district practice.
- The Design Team contributed significantly to the progress of the pilot.
- District support systems were seriously challenged by the implementation of Pay for Performance. Many opportunities for change were identified and district action resulted. Challenges of organizational alignment still lie ahead for the district.
- The turnover in leadership positions during the course of the pilot, particularly at the level of the pilot school principals and the superintendency, contributed to some of the concerns related to trust and institutional priority that have affected the implementation of the pilot.
- The lack of an agreed-upon and aligned portfolio of district assessments for measuring student achievement meant that 166 identifiable assessments were used to measure progress in meeting objectives, and 256 teachers used generally referenced measurements, in the last year of the pilot.
- The task of linking student achievement results to specific teachers has proven more challenging than originally anticipated by the district. As pilot efforts go to a broader scale of implementation in the district, this type of data capacity will be greatly needed.
- Several factors, including the state and national high stakes testing environment and the district's experiences with pay for performance for administrators, adversely affected the climate for implementing the pilot.

Charlotte

Executive Summary

Teacher Incentive Fund—Leadership for Educators’ Advanced Performance (TIF-LEAP), a multi-year performance-based compensation initiative in Charlotte-Mecklenburg Schools (CMS), focused on improving teaching and learning in a select group of high need schools.

Benefiting from a community culture that supports using monetary incentives to encourage and reward employee performance, the district partnered with the Community Training and Assistance Center (CTAC) in 2007 to seek, obtain and implement a Teacher Incentive Fund (TIF) grant from the U.S. Department of Education.

The TIF-LEAP initiative introduced two approaches to performance-based compensation—Student Learning Objectives (SLOs) and a value-added measure (VAM)—into a total of twenty schools over the course of implementation. The initiative established a structure through which teachers and principals earned bonuses for demonstrated increases in student academic growth.

Already complex and multi-layered, the initiative was seriously affected in years four and five as the recession roiled the district budget and accordingly district schools. CMS maintained its matching fiscal commitment to the TIF-LEAP initiative, but planning for and implementing teacher layoffs, principal changes, school closures, and school reorganizations influenced outcomes in the final years of implementation.

After a peak performance year (2009-10), with all twenty schools phased in and both approaches implemented, the following year (2010-11) turned into what evaluators came to call “the perfect storm.” Misunderstandings and misgivings about the VAM rankings among many participants and the implementation of a new teacher appraisal system started the year, while layoff notices and plans for closing and reorganizing selected schools dominated the spring. The TIF-LEAP schools experienced significant turnover in principals, and the superintendent left later in the spring. At the beginning of year five, the overall number of participating schools was reduced to eleven. The TIF-LEAP initiative concluded with the 2011-12 academic year.

It's More Than Money is the evaluation of the initiative, based on five years of observations, annual stakeholder surveys and interviews, and analyses of SLO artifacts and student achievement results. As the title of the report indicates, improving teaching and learning through performance-based compensation is an enterprise that does not run on the promise of monetary incentives alone. Success depends on more than money.

Promising Results

Student Academic Growth on North Carolina Assessments

The descriptive statistical analysis shows increases in student achievement attributable to the TIF-LEAP initiative. This analysis examines the North Carolina End-of-Grade student results from TIF-LEAP schools together with those of comparison schools:

- The growth rate of students in TIF-LEAP schools is greater than that of students in the comparison schools. Although the TIF-LEAP schools start with lower student performance, by the end of year four, the student test scores in both mathematics and reading are closely approaching those of the comparison schools.
- The TIF-LEAP schools show greater resilience to the negative shocks resulting from the economic recession, including teacher layoffs, and planning for school closures and restructuring that occurred in 2010-11. Student test scores in the TIF-LEAP schools grew at a lower rate in that year than in the previous school year. However, they grew at a higher rate than the comparison schools that experienced the same disruptions.

The longitudinal hierarchical linear models provide the estimated effects of the TIF-LEAP initiative on student achievement. They show that TIF-LEAP had a positive impact on the participating schools which is both statistically and practically significant. Specifically,

- In terms of mathematics achievement, students in TIF-LEAP schools on average have a growth rate 12% greater than students in the comparison schools.

This growth difference is substantial and means that the TIF-LEAP students are growing 12% more than the 0.8% annual growth rate of the comparison schools. This growth translates into 0.34 points annual growth difference between TIF-LEAP and comparison students. As a result, at the end of year four of the initiative, the test scores of students in the TIF-LEAP schools improved, cumulatively, 1.4 points more than students in the comparison schools. This growth brings the TIF-LEAP schools close to par with the comparison schools (students in TIF-LEAP schools started 1.5 points lower than students in the comparison schools at the beginning of the initiative).

- In terms of reading achievement, students in TIF-LEAP schools on average have a growth rate 13% greater than students in the comparison schools.

This growth difference is substantial and translates into 0.44 points annual growth difference between TIF-LEAP and comparison students. As a result, at the end of year four of the initiative, the test scores of students in the TIF-LEAP schools are only 0.7 points lower than those in the comparison schools. The initial test scores of the TIF-LEAP students started 2.5 points lower than students in the comparison schools.

Three cross-sectional HLM analyses were conducted over the course of the TIF-LEAP initiative. The findings of the cross-sectional HLM models vary by subject and year. The first cross-sectional analysis is for 2008-09, the first year of SLO implementation. The full SLO effects on student achievement were expected to phase in over several years of implementation. The findings in the first year support this expectation:

- There are positive, statistically significant associations between the attainment of Target SLOs and student achievement both in mathematics and in reading.
- There is no statistically significant association between the quality of SLOs (as indicated by the rubric rating) and student achievement in this first year.

The second cross-sectional analysis is for 2009-10. In terms of achieving higher student performance, this is the peak year of SLO implementation. The key findings are:

- There are positive, statistically significant associations between the quality of SLOs and student achievement. This finding means that a teacher's SLO rating relates positively to student achievement in elementary school mathematics, elementary school reading, and middle school mathematics.
- There are positive, statistically significant associations between the attainment of SLOs and student achievement at the elementary school level. This finding means that the students whose teachers met their SLOs achieved higher scores in elementary school mathematics and reading.

The third cross-sectional analysis is for 2010-11. In this school year, as a result of the increase in the number of students and classrooms, the investigation is conducted at the individual grade level in grades 4-8 rather than combining grades into elementary and middle school analyses. The key findings are:

- There is a positive, statistically significant association between the quality of SLOs and student achievement in mathematics in grade five.
- There is a positive, statistically significant association between the attainment of SLOs and student achievement in reading in grade six.

Student Learning Objectives

The district began implementing Student Learning Objectives in year two of the initiative as an approach to measuring the connection between teacher performance and student achievement. Specifically, SLOs were implemented in the first ten schools in 2008-09 and in ten additional schools which joined the initiative in 2009-10.

Nearly 4,000 teacher-developed SLOs were evaluated for this study, using CTAC's four-level rubric that examined content, expectations, completeness, and coherence. The study also examined (1) whether or not the growth targets set by teachers for their SLOs were attained (met or not met) and a bonus collected, and (2) the relationship between the quality of each SLO and the attainment of the growth target. The analysis showed:

- The overall relationship between the quality of SLOs and their attainment is positive. Year-by-year findings vary with the highest correlation found in 2009-10. The relationship between the quality of an SLO and its attainment (meeting or exceeding the growth target set by the teacher) is statistically significant. It shows that the higher the quality of the SLO, the greater the likelihood it will be attained.

Further, the number of years a teacher participates in SLO implementation matters.

- Teachers in the initiative for three years of SLO implementation develop higher quality SLOs and have greater success in attaining their SLOs. The relationships between the quality of an SLO and its attainment to the teacher's length of time in the initiative are statistically significant.

Setting and reaching SLO targets was a new practice for teachers and principals, and the learning curve for both participants and initiative staff is evident as teachers and TIF-LEAP staff gained experience between the first and second years of SLO implementation. The second year of SLO implementation (2009-10) is the strongest year overall for SLO-related performance, in particular, as well as for the fully implemented initiative, in general. The analysis of the third year of SLO implementation (year four of the initiative) finds decreases in some areas, as many teachers and schools were affected by the fiscal crisis which led to layoffs and school closures.

Survey and interview responses show:

- The Curriculum and Instruction Department identified SLOs as a district instructional best practice. Numerous teachers in the TIF-LEAP schools declared an intention to continue the process after the conclusion of the initiative.
- Beyond the bonus payouts, teachers valued the data analysis, planning, and instructional elements of the SLO process. These elements resonated with teachers as being significant to teaching students more effectively and to advancing their professional growth.
- The SLO process provided teachers and principals with the tools to look carefully at their students through the lens of more timely and better baseline data. They observe that they had greater capacity to analyze each student's progress and set growth targets that stretched and encouraged every student.
- The SLO baseline data analysis—prior to setting targets and planning instruction—prompted more in-depth analysis about the best instructional strategies to meet student needs.
- The new North Carolina Teacher Evaluation Process, implemented in year four of the initiative, reinforces the TIF-LEAP work with SLOs, according to teachers, principals, and the TIF-LEAP team.

Maryland

Executive Summary

The Maryland State Department of Education (MSDE) is guiding and supporting the implementation of a new Teacher and Principal Evaluation (TPE) system in all school districts throughout the state. The system includes measures of both professional practice and student growth. Because the historical and current practice in Maryland is one of local control, the key TPE implementation decisions and the organizational supports that reinforce those decisions are made at the district level.

This study examines the perceptions of frontline educators in Maryland of the support they receive in understanding and implementing the TPE system. It particularly focuses on the key component of Student Learning Objectives (SLOs), an evidence-based instructional process for improving educator effectiveness and student achievement. Findings of this study are intended to help further inform and strengthen the implementation of the overall TPE system and the SLO component, in particular, in Maryland.

Data for this study include interviews of leaders from 12 districts and the state level education association; a case study of four districts; and statewide survey responses provided by teachers, principals and other educators from 24 districts.

Findings show that TPE implementation is generating changes in practice and perception.

Findings

Overall Perceptions of TPE

- With each year of implementation, perceptions of both principals and teachers are increasingly more positive about TPE. As educators have more experience with implementing TPE, they are more familiar and comfortable with all aspects of the TPE processes.
- While both principals and teachers are positive about TPE, principals are generally more positive than teachers, but differences in opinion are narrowing.
- Maryland has successfully addressed the initial challenges of TPE. In the context of this progress, gaps remain.

School, District and State Support

- Survey results show more principals and teachers agree in 2015 than in 2014 that school, district, and state support for the implementation of TPE has been helpful.
- While the pattern of perceptions is increasingly more positive, these data indicate further TPE-related support is still needed to reach all frontline educators.

Instructional Dialogue

- Many educators indicate that principals and teachers are deepening their reflection, and having more data-driven and focused conversations on instruction in their school or district.

Implementing Observations

- In interviews and focus groups, Maryland's frontline educators indicate that actual observation practice often differs from best practice.
- Eighty percent of principals and 52% of teachers agree that the evaluation frameworks and processes use validated observation measures/instruments.
- Ninety-three percent of principals and 66% of teachers agree that observations are conducted by observers/evaluators qualified to do the evaluation.

Teacher Ratings

- In the districts where TPE is implemented from instructional and capacity building perspectives, educators see more value of the evaluation results in promoting instructional conversations and measuring teacher performance.
- When the implementation of TPE is driven to a greater extent by complying with mandates, there is generally more stress for teachers related to achieving the “highly effective” rating and a greater sense of loss of the reflection dimension of evaluation.
- Teachers are generally becoming more confident that the evaluation results will be used to inform their professional development next year.

The Effect of PARCC

- Effectively implementing a new evaluation system, and new standards, curricula and assessments all require substantial input of resources, particularly time, from teachers and principals. In interviews, educators express anxiety about the concurrent rollout of these initiatives.
- Educators have concerns on the many tests that students have to take. Many feel they lose too many instructional hours due to the assessments, a concern that is generally exacerbated when a major new assessment is introduced.

Key Issues in Implementing TPE

- Pivotal and inter-related concerns affecting implementation in all of Maryland’s districts are the quality of the TPE process, the consistency of quality within and across schools and districts, and the manageability of the new TPE systems.
- Emphasis on instruction versus compliance varies by district throughout the state. Districts generally appear in one of three groups: districts that are approaching TPE from an instructional perspective; districts with a range of capacities and implementation approaches; and districts that are implementing TPE because it is a state requirement.
- An ongoing challenge for Maryland districts and schools is to integrate a number of initiatives into a strategically coherent instructional improvement process.

SLO Quality Matters

- Districts are beginning to make progress in providing information to teachers about what constitutes a high quality SLO, but teachers still need more customized professional development on all aspects of SLOs to strengthen SLO quality and improve the consistency of that quality.
- Educators report that several districts are also making improvements in the quality and inter-rater reliability of SLO ratings.

Quality in the Overall SLO Process

- Maryland is making continued progress in preparing educators to implement the SLO process with greater fidelity. Over the past three years of implementation, survey and interview data indicate more educators agree that teachers are receiving support and resources on the SLO process; however, gaps remain.
- Teachers still want and need additional SLO-related professional development to improve their ability to select learning content, identify research-based instructional strategies, and set growth targets.

Broader Learnings about Support

- In districts that make a conscientious effort to improve the SLO process, teachers and administrators agree in interviews that the SLO process becomes easier and more effective with experience.
- Many interviewees feel that there is less fear and more understanding about the TPE process, and they are challenging themselves to develop better SLOs, and as a consequence having better dialogue and results.

Teacher Capacity and Support

- More teachers in 2015 compared to 2014 say they are receiving the support they need to implement TPE. Despite these improvements, approximately 45% of teachers indicate that they need further training.
- Teachers have needs in areas in which principals indicate similar and often greater need.

Principal Capacity and Support

- Principals' unmet needs in key capacity building areas affect their ability to support and guide their respective school staffs in the same areas.
- One area where interviewees feel principals need additional professional development is in providing feedback and leading a constructive dialogue on instruction.
- Principals want to see districts pay more attention to principal evaluation, giving principals the opportunity for meaningful feedback and dialogue around their instructional leadership.

Communication and Collaboration

- Where the teacher/district communication is two-way and considered genuine and respectful, interviewees indicate that implementation is more effective.
- In several districts, the quality of communication is also contributing to a higher level of union/district collaboration.
- Where the communication is perceived as being top-down, one way, or giving more the appearance than the reality of partnership, there is less buy-in from school site practitioners and TPE is often perceived as a compliance activity.

Common Language

- A factor that is enhancing the implementation of the new evaluation system—or key components of that system—is when district and school level practitioners develop a common language about instruction.
- In every district, more teachers and principals agree than disagree that they have a common language to describe the SLO process, and that expectations are clear.

Managing Technology and Using Data

- Districts are generally not yet making systematic use of the information generated through TPE. Interviewees describe technology systems that serve primarily as a repository of information, rather than as a management system.
- Districts need to examine their technology tools to understand how and to what extent principals and teachers make use of them.

Learnings from a Case Study of Four Districts

- When implemented with an instructional focus, attention to quality and an organizational commitment to achieving greater consistency of that quality, TPE promotes positive change in principal and teacher practice.
- When these factors are lacking, implementation is markedly less effective.